

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS QUÍMICAS
DEPARTAMENTO DE QUÍMICA FÍSICA I



TESIS DOCTORAL

Modelos simples para estudiar el efecto de la presión en el plegamiento de proteínas

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Ramiro Perezzan Rodríguez

Director

Antonio Rey Gayo

Madrid, 2015

Universidad Complutense de Madrid
Facultad de Ciencias Químicas
Departamento de Química Física I



**Modelos simples para
estudiar el efecto de la presión
en el plegamiento de proteínas**

Memoria para optar
al grado de Doctor por la Universidad
Complutense de Madrid
realizada por

Ramiro Perezzan Rodriguez

Director:
Prof. Antonio Rey Gayo

Dpto. Química-Física I

Madrid, 2014

Índice general

1. Introducción	1
1.1. Las proteínas	2
1.1.1. De la química a la biología de las proteínas	3
1.1.2. El plegamiento de proteínas	9
1.1.3. Fuerzas que dirigen el plegamiento de proteínas	12
1.1.4. Tipos de transiciones en el plegamiento	13
1.2. La presión	15
1.2.1. Presión hidrostática	16
1.2.2. Aplicaciones de la presión como magnitud termodinámica	17
1.3. Influencia de la presión en el plegamiento de proteínas	19
1.3.1. El diagrama de fases p - T de las proteínas	20
1.3.2. Diferentes mecanismos de desnaturalización	22
1.3.3. Factores que afectan a la desnaturalización mediada por la presión	25
1.3.4. Aplicaciones de la desnaturalización mediada por presión	27
1.4. Estudio del plegamiento de proteínas	28
1.5. Objetivos y organización de la Tesis	32
2. Materiales y métodos	35
2.1. Motivación del Capítulo	37
2.2. Modelo geométrico	41

2.2.1. Potencial de interacción	44
2.2.2. Mapa de contactos	51
2.3. Método de simulación	55
2.4. Análisis de Datos	62
3. Aplicación del potencial de alta presión	69
3.1. Motivación del Capítulo	70
3.2. Selección de la proteína	71
3.2.1. Mapa de contactos de la proteína 2GB1	74
3.3. Calibrado del potencial de interacción	76
3.4. Herramientas para el estudio del tipo de plegamiento	79
3.4.1. La curva de capacidad calorífica	79
3.4.2. El perfil de energía de Helmholtz	83
3.4.3. Histogramas y representaciones bidimensionales	85
3.5. El tránsito para la proteína 2GB1	87
3.5.1. Análisis del incremento de presión	87
3.5.2. Evaluación de los diferentes estados involucrados en la transición .	95
3.6. Estudio de poblaciones de los diferentes estados	106
3.6.1. Diagrama de fases $\epsilon_{ssm}-T$ con nuestro modelo	113
3.7. Análisis a nivel local de las diferentes conformaciones	117
3.7.1. El estado nativo y el estado desnaturalizado	120
3.7.2. Comportamiento estructural a alta presión	122
3.7.3. El estado hinchado	126
3.7.4. Evolución del estado hinchado con la presión.	128
3.8. Resumen y conclusiones del Capítulo	130
4. Evaluación del modelo en proteínas con distinta estructura secundaria	135
4.1. Motivación del Capítulo	137

4.2. Proteínas con un solo tipo de estructura secundaria	139
4.2.1. Proteínas α	139
4.2.2. Proteínas β	140
4.2.3. Diferencias entre las dos familias de proteínas	142
4.3. Resultados al aplicar nuestro potencial de interacción	145
4.3.1. Tipos de plegamiento observado	146
4.3.2. Evolución de los diferentes estados	151
4.3.3. Análisis de los estados hinchados	155
4.3.4. El diagramas de fases p - T para los cuatro sistemas	163
4.4. Resumen y conclusiones del Capítulo	165
5. Comparación con resultados experimentales	169
5.1. Motivación del Capítulo	169
5.2. La proteína SNasa	171
5.3. El plegamiento a presión atmosférica	173
5.4. Aparición del estado hinchado con el aumento de presión.	176
5.5. Comparación de los diagramas de fases.	183
5.6. Resumen y conclusiones del Capítulo.	188
6. Resumen y conclusiones generales de esta Tesis	191
7. Summary	195
7.1. Introduction.	195
7.2. Main Objectives.	197
7.3. Methods.	197
7.4. Results.	199
7.5. Conclusions.	201

Capítulo 1

Introducción

Los seres vivos son máquinas de la naturaleza perfectamente ensambladas para poder adaptarse y evolucionar en el medio en el que se encuentran. Todos los seres vivos a su vez están compuestos por millones de moléculas y la interacción de estas moléculas individuales entre sí es la que permite la vida a las diferentes especies. Las proteínas son, junto a los ácidos nucleicos, carbohidratos y lípidos, uno de los cuatro tipos de moléculas biológicas más importantes que existen en los seres vivos¹. La importancia de las proteínas radica en la gran variedad y especificidad de funciones que pueden realizar en el interior de los organismos vivos, ya que procesos tan dispares como el transporte a través de la membrana celular, el transporte de oxígeno en la sangre, el movimiento muscular o la respuesta inmune del organismo, son todos procesos regulados por proteínas, y necesarios para la existencia de la vida tal y como la conocemos.

Las proteínas actúan como pequeñas máquinas biológicas capaces de realizar multitud de reacciones químicas, en muchos casos muy específicas. Un subconjunto de las proteínas, las denominadas enzimas, son catalizadores biológicos altamente específicos que tienen la capacidad de realizar una reacción concreta en el momento adecuado a pesar de encontrarse en el interior de los organismos, un medio que desde el punto de vista de un químico podría parecer especialmente complejo para llevar a cabo ciertas reacciones. Por tanto, como cualquier sustancia capaz de reaccionar, estas enzimas de-

penderán de magnitudes físicas que pueden afectar en su reactividad, como por ejemplo la temperatura, el pH o la presión. El poder conocer cómo cualquiera de estas propiedades afecta a los organismos vivos y específicamente a las moléculas que los componen resulta ser de especial interés de cara a conocer cómo y en qué manera los seres vivos regulan el interior de su organismo para dar lugar a la vida.

1.1. Las proteínas

Las proteínas debido a su alto peso molecular son macromoléculas que forman parte de los seres vivos sea cual sea su complejidad, desde organismos unicelulares hasta el mamífero de mayor tamaño. Desde un punto de vista biológico existen distintas familias de proteínas que tienen diferentes finalidades dentro de los organismos vivos. Así, podemos encontrar **proteínas reguladoras**, como la insulina o las hormonas del crecimiento, **proteínas de transporte**, como la hemoglobina y la mioglobina, **proteínas de defensa en los vertebrados**, como son los anticuerpos, **proteínas de membrana**, ejemplo de este grupo son los receptores de membrana y los antígenos, **proteínas con carácter enzimático**, como por ejemplo la sacarasa y la pepsina, **proteínas de reserva de energía**, es el caso de las albúminas, un tipo de proteína muy frecuente en la clara de huevo o de muchas proteínas presentes en las semillas, o **proteínas estructurales**, como el colágeno, la seda o la queratina¹. Estas son sólo algunas de las familias de proteínas más importantes en los seres vivos.

Pese a existir un inmenso número de proteínas, todas ellas comparten una característica principal, su estructura, ya que pese a ser moléculas flexibles de gran tamaño, muchas de ellas tienen la capacidad de adquirir una conformación específica, y sólo bajo esa conformación son biológicamente activas. A la conformación de la proteína biológicamente activa se la denomina **estado nativo** de la proteína. Cuando pierden esa conformación adecuada, su actividad biológica disminuye considerablemente, llegando en muchos casos a hacerse despreciable o incluso nula. Precisamente, la versatilidad fun-

cional existente en las proteínas depende de la configuración que adoptan y las moléculas con las que interaccionan. Esta característica ha sido utilizada a lo largo de la evolución por los diferentes organismos para regular la mayoría de las reacciones que se dan en el interior de los seres vivos. Por esta razón, se sabe que **existe una relación directa entre la estructura de las proteínas y su función biológica**.

Existen tres tipos de familias de proteínas si atendemos a su estructura²: **proteínas fibrilares**, cuya función principal es de tipo estructural, **proteínas de membrana**, encargadas del transporte de moléculas a través de las membranas celulares, y **proteínas globulares**. Estas últimas son las más abundantes y las encargadas de realizar la mayoría de las reacciones en el interior y exterior de la célula. En este trabajo trataremos con esta última familia de proteínas que son solubles en agua, tienen una tendencia a formar estructuras compactas y son las de mayor variabilidad en cuanto a su secuencia y estructura.

1.1.1. De la química a la biología de las proteínas

Todas las proteínas están compuesta por una sucesión lineal de aminoácidos. Desde un punto de vista químico, se puede considerar a las proteínas como un copolímero donde cada uno de los monómeros que lo componen corresponde a un aminoácido. Los aminoácidos son moléculas orgánicas sencillas, que como su nombre indica contienen un grupo amino ($-NH_2$) de naturaleza básica, y un grupo carboxílico ($-COOH$) de naturaleza ácida. Ambos grupos se encuentran situados en el mismo átomo de carbono (denominado carbono α), por esta razón se les denomina α -aminoácidos.

Existen 20 aminoácidos naturales proteicos que se caracterizan por tener cada uno su propia cadena lateral ($-R$), que también está unida al carbono α , de tal forma que, como se puede ver en la Figura 1.1, este átomo de carbono se encuentra unido a cuatro sustituyentes distintos, los tres ya mencionados y un átomo de hidrógeno. Esto le confiere isomería óptica³. De los dos diastereoisómeros posibles, sólo el L está presente en

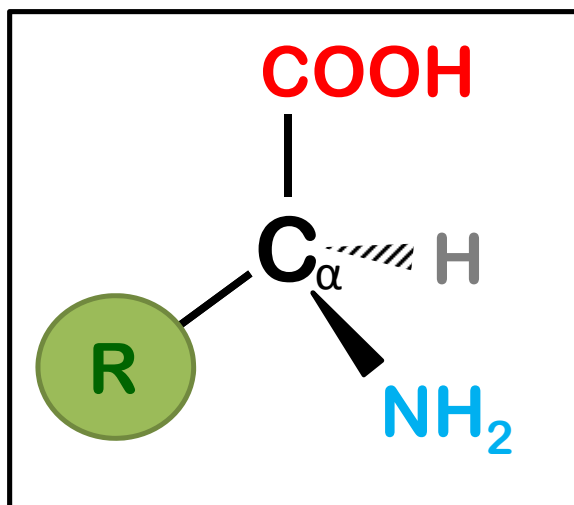


Figura 1.1: **Estructura de un α -aminoácido.** El carbono α del aminoácido soporta cuatro sustituyentes diferentes, lo que le confiere propiedades enantioméricas. El grupo R corresponde a la cadena lateral, específica de cada aminoácido.

los aminoácidos naturales⁴. Por tanto, es únicamente la cadena lateral la que diferencia a cada uno de los 20 aminoácidos naturales. Además, la cadena lateral confiere a cada aminoácido unas propiedades específicas en función de los átomos que la componen. La clasificación más usual para los aminoácidos naturales, que se puede ver en la Figura 1.2, responde precisamente a las características de las cadenas laterales. Los 20 aminoácidos naturales se pueden dividir en los 4 grupos siguientes:

1. **Aminoácidos apolares.** Se trata de aminoácidos cuya cadena lateral es de carácter hidrófobo. Como se puede ver en la Figura 1.2, este grupo de aminoácidos es el más grande en número. Se caracterizan por presentar una tendencia a agruparse entre sí en medio acuoso a través de interacciones hidrófobas, confiriendo mucha estabilidad a la estructura nativa de la proteína. En realidad no existe una fuerza atractiva entre residuos apolares, sino que se obtiene una mayor estabilidad del sistema al minimizar el número de moléculas de agua necesarias para rodear fragmentos hidrófobos expuestos de la proteína⁵. Dentro de este grupo, además de los aminoácidos con carácter hidrófobo se incluye a la glicina, que tiene como único sustituyente en la cadena lateral un átomo de hidrógeno, y se trata por tanto del

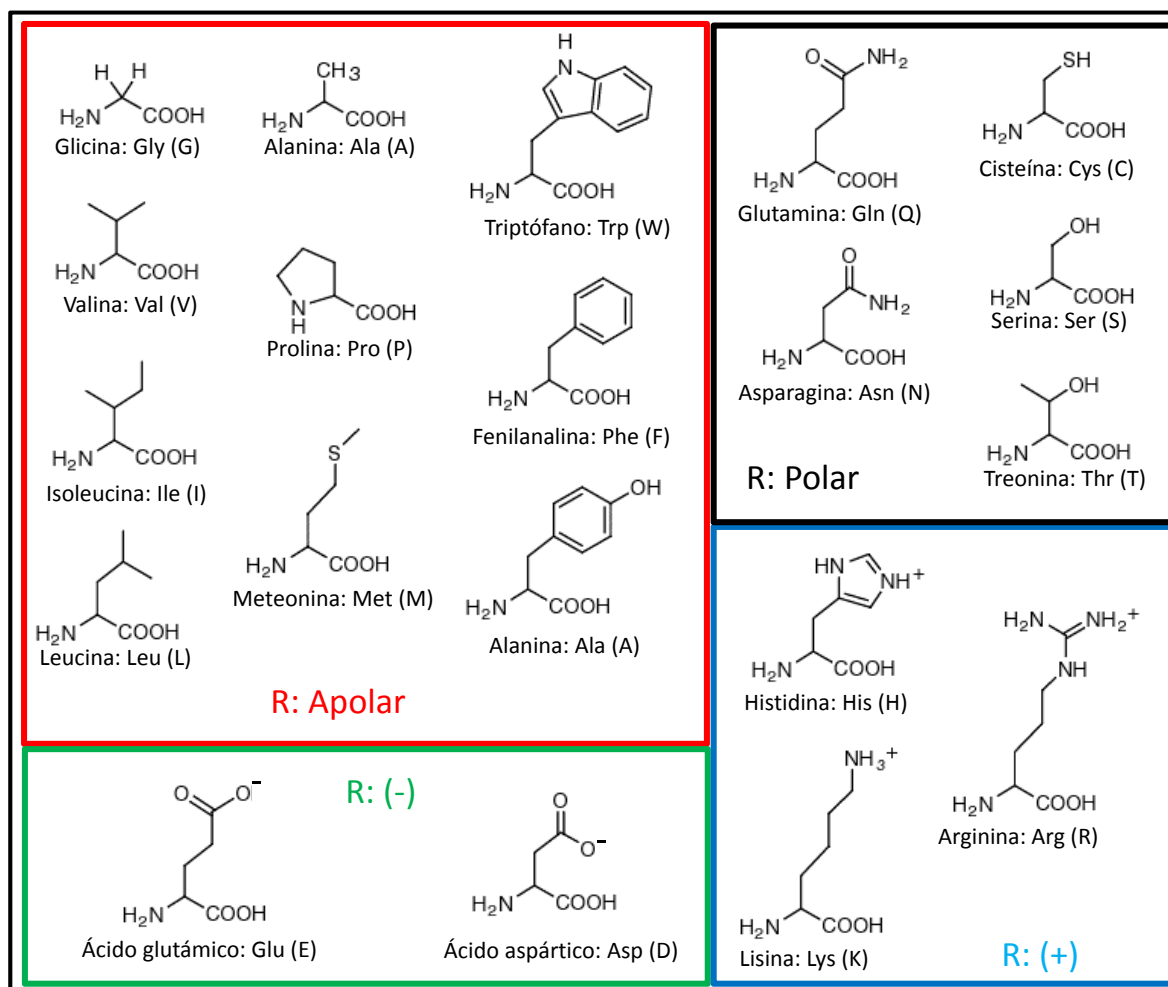


Figura 1.2: **Clasificación de los aminoácidos.** Representación de todos los aminoácidos naturales clasificados en función de las características de su cadena lateral.

único aminoácido que no presenta isomería óptica. Además, también están incluidos la tirosina y el triptófano que, siendo aminoácidos con anillos aromáticos y por tanto de carácter hidrófobo, presentan a la vez grupos polares en sus cadenas laterales.

2. **Aminoácidos polares.** Son aminoácidos sin carga neta pero que presentan carácter hidrófilo, debido a la presencia de un grupo alcohol ($-OH$), un grupo tiol ($-SH$) o un grupo amida ($-COONH_2$) en su cadena lateral. Lo interesante de los aminoácidos polares es que sus cadenas laterales tienen la posibilidad de formar enlaces de hidrógeno entre sí o con el disolvente, lo cual servirá para dar mayor

estabilidad a la proteína en disolución.

3. **Aminoácidos cargados positivamente.** Nos encontramos ante aminoácidos con cadenas laterales que están cargadas positivamente a pH neutro. Estos presentan grupos con átomos de nitrógeno en la cadena lateral, que a pH neutro se encuentran protonados. Esto les confiere un carácter muy polar y se suelen situar en la parte exterior de las proteínas, donde pueden hidratarse en el medio acuoso que les rodea.
4. **Aminoácidos cargados negativamente.** Existen dos aminoácidos que a pH neutro se encuentran cargados negativamente, debido a la presencia de un grupo carboxilo desprotonado. En este caso esa carga negativa les confiere un carácter muy polar, y de nuevo les llevará a disponerse en la superficie de la proteína.

Estos 20 tipos de aminoácidos están unidos en las proteínas a través de un enlace característico, **el enlace peptídico**, para dar lugar a una cadena polipeptídica. El enlace peptídico se forma entre el grupo α -carboxilo de un aminoácido y el grupo α -amino del siguiente aminoácido, a través de una reacción de condensación que da lugar a la formación de la amida correspondiente y a la liberación de una molécula de agua⁶. El producto formado de esta unión se denomina péptido. El enlace peptídico, al tratarse del enlace propio de una amida, presenta un carácter parcial de doble enlace, ya que presenta dos estructuras en resonancia entre sí, lo que le confiere algunas características importantes: se trata de un enlace donde la rotación alrededor de su eje esta impedida, lo que lo convierte en un enlace con isomería de tipo *cis/trans*. Energéticamente está favorecido el isómero *trans* al presentar menor impedimento estérico que la forma *cis*. Además, esta rigidez en el enlace peptídico hace que la flexibilidad del biopolímero esté considerablemente reducida y limita su libertad conformacional. En la Figura 1.3 mostramos dos enlaces peptídicos de un mismo aminoácido. Se puede observar cómo el resto de átomos unidos a los dos que forman el enlace peptídico se encuentran en

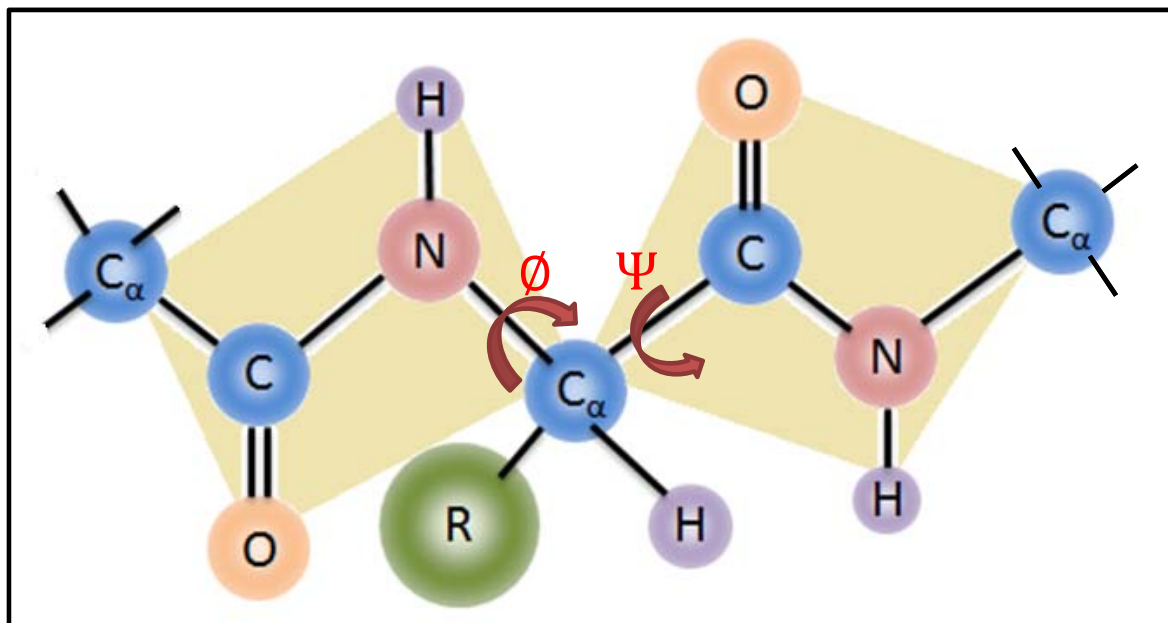


Figura 1.3: **Restricciones alrededor del enlace peptídico.** El C_{α} situado entre dos enlaces peptídicos pertenece simultáneamente a dos planos diferentes que podrán girar en torno a los enlaces ϕ y ψ . Sin embargo, el giro sobre estos dos enlaces está restringido por el impedimento estérico que genera la cadena lateral (R). Imagen adaptada de⁷.

un mismo plano, coloreado en amarillo, y cómo alrededor del C_{α} central los átomos se disponen en dos planos que pueden girar alrededor de los ángulos ϕ y ψ .

La libertad conformacional se ve aún más restringida debido a la presencia de la cadena lateral (R), que genera impedimento estérico con el resto de la cadena polipeptídica impidiendo que los ángulos ϕ y ψ giren libremente a pesar de ser sencillos⁷. Los valores que toman estos dos ángulos están restringidos a intervalos concretos, lo que se corresponde con ciertas conformaciones específicas. Por eso la flexibilidad de cada cadena lateral es característica de la propia cadena lateral y de los aminoácidos cercanos. En cualquier caso, las cadenas laterales no se disponen al azar en la estructura nativa de una proteína, sino que se disponen de forma compacta creando un buen empaquetamiento alrededor de la cadena principal⁸. Si bien pudiese parecer que el número de conformaciones posibles se ve muy limitado debido a las distintas restricciones geométricas, esto no es verdad ya que existen multitud de conformaciones para cada cadena polipeptídica, lo que hace si cabe más sorprendente, en muchas de ellas, la necesidad de que adopten

una estructura tridimensional única para poder ser biológicamente activas.

Precisamente, para que la cadena polipeptídica pueda adoptar la conformación adecuada para realizar su función, las proteínas deben adoptar distintos niveles de complejidad en su estructura. A la cadena simple de aminoácidos unidos mediante enlaces peptídicos se la denomina **estructura primaria** de las proteínas. La cual recoge mucha información acerca de la funcionalidad de la proteína, ya que la secuencia de aminoácidos en la cadena será la que dictamine la estructura tridimensional de la proteína. La cadena de aminoácidos a su vez adopta conformaciones que estabilizan a la proteína, guiada por las restricciones propias del enlace peptídico y de las cadenas laterales. Frecuentemente lo hace de tal forma que los ángulos ϕ y ψ toman valores regulares en varios aminoácidos consecutivos que nos llevan a que adopten situaciones de ordenamiento local, dando lugar a ciertos motivos estructurales que se repiten en muchas proteínas. La ordenación a nivel local de la cadena constituye la **estructura secundaria** de la proteínas. Los elementos de estructura secundaria más importantes son las **hélices** α y las **láminas** β , que son estructuras con una forma determinada y una direccionalidad específica como se puede ver en la Figura 1.4. La ordenación espacial de los diferentes elementos de estructura secundaria es lo que se conoce como la **estructura terciaria** de las proteínas, que nos lleva a estructuras compactas. Existen muchas proteínas que necesitan más de una cadena de aminoácidos para ser funcionales. La unión de dos o más sub-unidades terciarias con entidad independiente para dar lugar a un único complejo constituye la **estructura cuaternaria** de las proteínas; se trata del nivel de estructura más complejo en las proteínas. En muchos casos, sólo cuando las proteínas se encuentran en su estructura terciaria o cuaternaria pueden ser biológicamente activas. En la Figura 1.4 se muestra una representación de los diferentes niveles de complejidad estructural que existen en las proteínas.

En definitiva, existen multitud de proteínas con funciones muy diversas, y la funcionalidad de cada proteína depende con frecuencia de su estructura. Por tanto, la

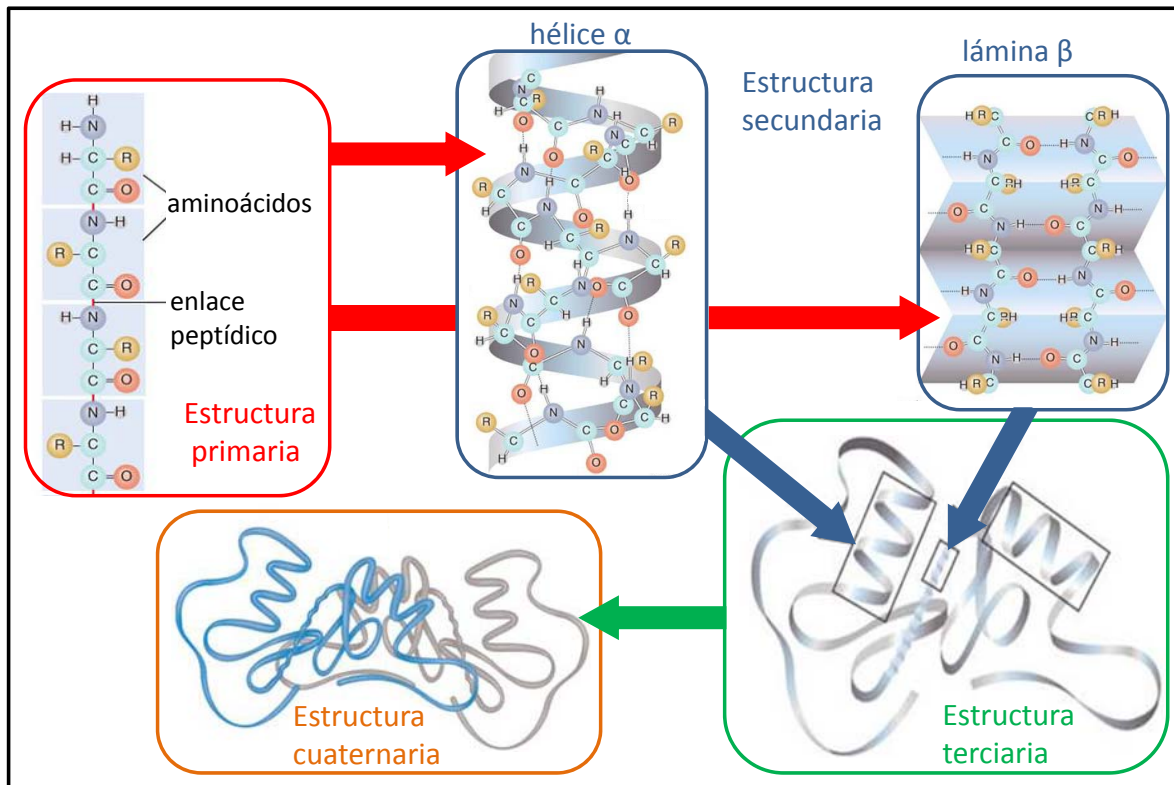


Figura 1.4: **Organización estructural de las proteínas.** Se muestran los cuatro niveles de complejidad estructural de las proteínas. 1º: secuencia covalente de aminoácidos, 2º: ordenación a nivel local, 3º: organización tridimensional de los elementos de estructura secundaria, 4º: asociación de más de una cadena independiente.

secuencia de aminoácidos que compone la cadena polipeptídica será, a través de diferentes tipos de interacciones que se establecen entre ellos, la que dictamine la estructura de la proteína, y por tanto también su función.

1.1.2. El plegamiento de proteínas

El proceso por el cual la proteína va adquiriendo una mayor complejidad estructural, pasando desde la estructura primaria a la estructuras terciaria o cuaternaria se denomina **plegamiento de la proteína**. Dado que se trata de un proceso reversible, cuando invertimos este tránsito y pasamos del estado nativo de la proteína a un estado que no es biológicamente activo, el proceso se denomina **desnaturalización de la proteína**, puesto que el estado no nativo de una proteína es conocido como **estado**

desnaturalizado. Este proceso, conocido e investigado desde hace décadas^{9,10}, no deja de ser sorprendente y espectacular para la comunidad científica, y a día de hoy prosigue con muchas preguntas sin respuesta. Resulta interesante poder conocer cómo es posible que con sólo la información en la secuencia de la cadena polipeptídica, la proteína sea capaz de plegar a un estado nativo altamente específico. Esta cuestión es conocida como el **problema del plegamiento**¹¹. Además, existe otro aspecto interesante acerca de este proceso, ya que se trata de un proceso que se realiza en un periodo de tiempo breve, en una escala de tiempos que va desde los milisegundos a los segundos, y de alguna forma este proceso tiene que ser guiado por algún tipo de fuerza directriz, puesto que un sistema con tantos grados de libertad como una proteína no podría llegar hasta el estado nativo, en tan poco tiempo, a través en un proceso puramente aleatorio¹².

Bajo un punto de vista termodinámico, el plegamiento de la proteínas puede ser considerado como una transición en el sistema donde el estado nativo se considera como el estado de menor energía en el sistema¹³. Se ha propuesto que la representación de la superficie de energía potencial del proceso de plegamiento tiene forma de embudo^{14,15}, como podemos observar en la Figura 1.5. La forma de este embudo es característica de cada proteína y del tipo de plegamiento¹⁶. La profundidad del embudo está relacionada con la energía del sistema, disminuyendo la energía a medida que el embudo se estrecha, de tal forma que el estado nativo se encuentra en la parte inferior del embudo. Al mismo tiempo, la anchura del embudo está relacionada con la entropía conformacional del sistema. Por esta razón, el embudo es más ancho en su parte superior, donde coexisten las diferentes conformaciones del estado desnaturalizado, de elevada entropía y degeneradas en energía, y se va estrechando a medida que nos acercamos al estado nativo, que se caracteriza por ser esencialmente único, por tanto mínimamente degenerado y de mínima entropía. Cada punto en la superficie de dicho embudo representa una de las configuraciones posibles del proceso de plegamiento. Como se puede ver en la Figura 1.5, la superficie de este embudo no es lisa, y serán precisamente las irregularidades en su

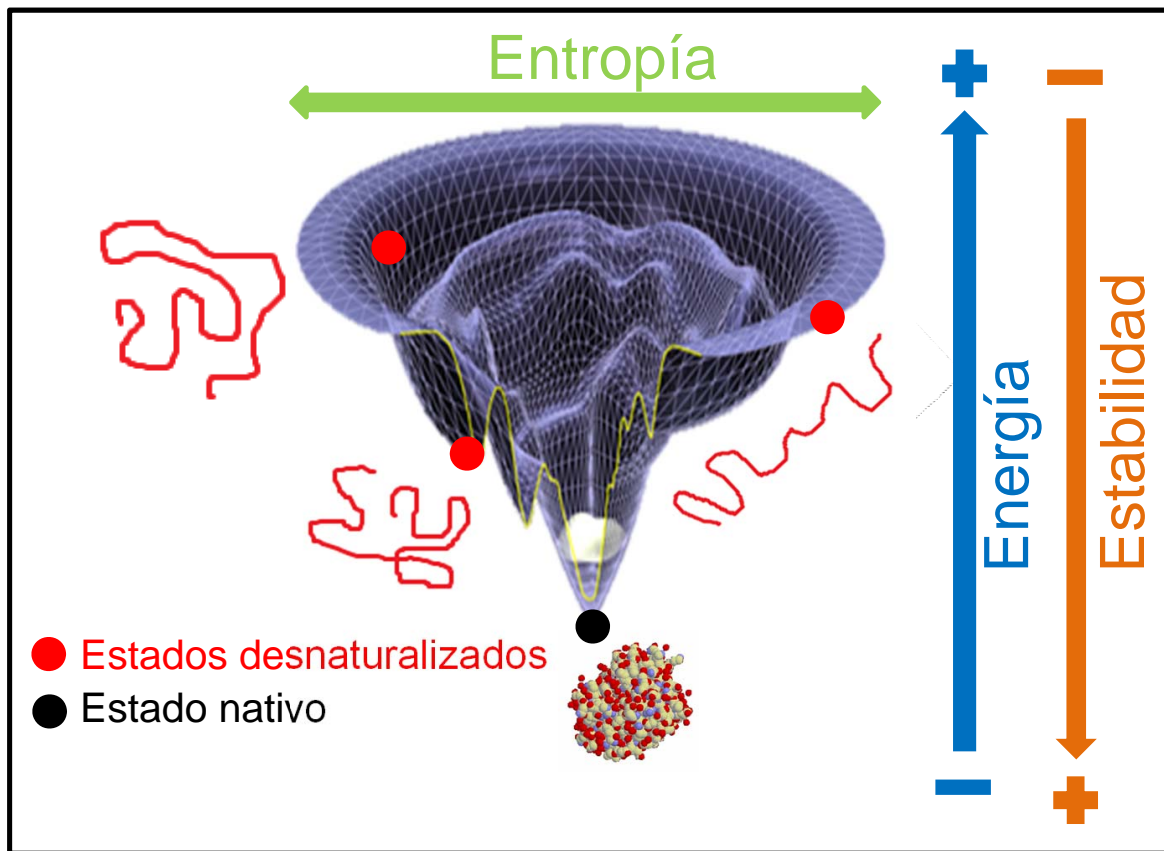


Figura 1.5: **Embudo de plegamiento de una proteína.** La representación de la superficie de energía potencial del proceso de plegamiento tiene forma de embudo, la altura del embudo está relacionada con la energía, y la anchura con la entropía del sistema. La superficie irregular será la que determine el camino de plegamiento.

superficie las que dirijan el camino que selecciona la proteína en su plegamiento, ya que desde cualquier conformación desnaturalizada se puede avanzar a lo largo de la superficie del embudo para llegar hasta la estructura nativa, obteniéndose un camino de plegamiento¹⁷. Dicho plegamiento por tanto es dependiente de las irregularidades de la superficie. A lo largo del proceso de plegamiento, existe una serie de interacciones entre los distintos aminoácidos que, de acuerdo con el principio de frustración mínima¹⁸, llevan a la proteína a adoptar el estado nativo.

La forma irregular de esta superficie de energía puede dar lugar a mínimos locales, donde la proteína estabiliza de manera transitoria algún estado desnaturalizado de mayor energía que el estado nativo^{19,20}. En caso de que esta configuración de la proteína fuese suficientemente estable, podría dar lugar a intermedios cinéticos o termodinámicos en el

proceso de plegamiento²¹. De alguna forma, ha sido la evolución la que ha seleccionado las distintas secuencias concretas de aminoácidos, las cuales, gracias a las fuerzas intra e intermoleculares que se establecen a lo largo del proceso, llevan de manera unívoca a una única estructura nativa a partir de la información secuencial de aminoácidos. Poder comprender a la perfección este proceso natural podría llevar a predecir de manera inequívoca la estructura tridimensional de un estado nativo a partir de su secuencia de aminoácidos, o por el contrario poder seleccionar qué secuencia de aminoácidos es necesaria para poder llegar hasta una estructura determinada, lo que resultaría ser de especial interés.

1.1.3. Fuerzas que dirigen el plegamiento de proteínas

Si bien hemos entendido que es la secuencia de aminoácidos la que dictamina la estructura nativa de la proteínas, hemos mencionado la existencia de interacciones que van dirigiendo la transformación del estado desnaturalizado hacia el estado nativo. Las interacciones responsables de la estabilidad del estado nativo son de distinta índole: participan **enlaces de hidrógeno**, que se pueden formar tanto entre aminoácidos como entre aminoácidos y las moléculas de disolvente acuoso en el que se encuentran las proteínas en su estado natural. Pero existen también **interacciones de tipo electrostático** entre grupos cargados, e **interacciones hidrófobas** entre los grupos hidrófobos de la cadena peptídica²². Aunque son muchas y de distinto tipo las interacciones que estabilizan el estado nativo, cabe destacar que con diferencia son las interacciones hidrófobas las principales responsables de dirigir el plegamiento²³. Estas interacciones se generan entre las cadenas laterales de los aminoácidos, de los que muchos de ellos como ya comentamos tienen un fuerte carácter hidrófobo. Por tanto, nos encontramos cómo de nuevo la selección de la secuencia va a ser determinante a la hora de conocer el camino que tomará la proteína hasta su estado nativo. Es tan importante el efecto hidrófobo en el plegamiento que existen teorías que indican que la estructura secundaria de la proteí-

na, principalmente estabilizada a su vez por enlaces de hidrógeno, es una consecuencia de la estructura terciaria, y que es el empaquetamiento hidrófobo el que permite un acercamiento entre residuos que cuando forman enlaces de hidrógeno intramoleculares estabilizan la estructura secundaria de una proteína^{22,24}. También existen experimentos que muestran cómo secuencias mutadas con distintos aminoácidos pero manteniendo los patrones de hidrofobicidad de la cadena, pliegan hacia la misma estructura nativa²⁵, lo que demuestra la importancia del efecto hidrófobo en el plegamiento proteico. Por tanto, la información que necesita la proteína para un correcto plegamiento se encuentra codificada en su secuencia de aminoácidos y en las interacciones que se forman entre los distintos aminoácidos y con el entorno, para así llegar de forma única a la estructura nativa.

1.1.4. Tipos de transiciones en el plegamiento

Aunque sea siempre la superficie del embudo la que dictamina el camino de plegamiento, el tránsito que existe entre el estado nativo de la proteína y el estado desnaturalizado puede ser de diferentes tipos. En un primer momento, utilizando el principio de frustración mínima, se describió el plegamiento de proteínas como un análogo a una transición termodinámica de primer orden¹⁸, y por tanto permitía describir el tránsito entre el estado nativo y el desnaturalizado como un equilibrio tipo ‘todo o nada’ o **equilibrio de dos estados**, donde a la temperatura de transición existe una coexistencia de ambos estados en equilibrio. Tenemos un estado de mayor energía, favorecido entrópicamente, que corresponde al estado desnaturalizado, y un estado de menor energía, favorecido entálpicamente, que corresponde al estado nativo de la proteína. Ambos estados se encuentran en equilibrio termodinámico durante la transición y a medida que aparece uno desaparece el otro. El proceso de dos estados ha sido ampliamente aceptado y existe la certeza de que muchas proteínas con un solo dominio siguen este tipo de transición en su plegamiento²⁶. La representación de las poblaciones de los diferentes

estados a la temperatura de transición en este tipo de tránsitos muestra la aparición de dos regiones separadas por una barrera de energía, que corresponde con una situación en la que no existen estados con población apreciable diferentes a los de partida y llegada.

Existen proteínas de mayor tamaño que pliegan en tres o más estados²⁷; para este tipo de tránsitos existe al menos un estado intermedio entre el estado nativo y el estado desnaturalizado de la proteína. En este caso a lo largo de la transición podemos encontrar un subconjunto del total de estados, o incluso la aparición simultánea de todos ellos dependiendo de la proteína estudiada²⁸. La representación de la población de estados a las diferentes temperaturas muestra, además de la aparición del estado nativo y del desnaturalizado, tantas regiones como intermedios existan entre el estado nativo y el estado desnaturalizado. El equilibrio de tres o más estados es, en realidad, una evolución del equilibrio de dos estados, y por tanto también puede ser considerado como varios equilibrios de tipo ‘todo o nada’.

Por último, en los últimos años se han caracterizado algunas proteínas de pequeño tamaño cuyo plegamiento no presenta características de tránsitos de dos estados²⁹, y son normalmente proteínas que siguen un plegamiento ultra-rápido. Para este tipo de plegamiento, denominado sin barrera o **transición tipo downhill**, el paso entre el estado nativo y el estado desnaturalizado no se da mediante un equilibrio, sino mediante una transición continua, donde el estado inicial evoluciona de forma progresiva a lo largo de la transición hasta alcanzar el estado final^{30,31}. De esta forma, el intervalo de temperaturas en el que se da el tránsito es más amplio que el intervalo de temperaturas en el que se da el equilibrio de dos estados. A lo largo del tránsito existe una evolución continua, sin llegar a observarse a ninguna temperatura la coexistencia de ambos estados. Cuando representamos la densidad de estados frente a la energía a una temperatura cercana a la de transición, lo que observamos es la presencia de un único pico en la representación, a una energía intermedia entre la del estado nativo y la del estado desnaturalizado. Esto confirma la presencia de un macroestado con propiedades intermedias entre las del

nativo y las del desnaturalizado.

Por tanto, se postulan al menos dos mecanismos diferentes para el plegamiento de proteínas, puesto que el equilibrio de tres o más estados se puede considerar una extensión del equilibrio de dos estados. Estos son el **equilibrio de dos estados** y el **tránsito tipo downhill**. El primero se trata de un equilibrio donde energéticamente existe una barrera entre el estado de partida y el de llegada y ausencia significativa de estados intermedios entre ambos estados, mientras que en el segundo caso se trata de un tránsito continuo donde el estado inicial evoluciona hasta el estado final de manera progresiva y adoptando configuraciones intermedias con población considerable entre ambos estados.

Hasta este momento hemos visto la estrecha relación que existe entre la estructura de estas macromoléculas y la función que ejercen en los seres vivos. Además, se ha explicado cómo la secuencia de aminoácidos en la cadena polipeptídica dictamina el camino de plegamiento que nos lleva desde el estado desnaturalizado hasta el estado nativo. Por último, hemos podido conocer los dos principales tipos de plegamiento que se describen en la bibliografía. A continuación mostraremos algunos aspectos generales de la variable principal que queremos introducir en nuestro sistema, la presión, para así poder conocer de manera completa el sistema que deseamos estudiar en esta Tesis.

1.2. La presión

Aunque la palabra presión sea utilizada habitualmente dentro de nuestro vocabulario común en varias de sus múltiples acepciones, desde un punto de vista científico la presión viene definida como una magnitud física que expresa la fuerza ejercida sobre un cuerpo por unidad de superficie. Por tanto, cuando sobre una superficie de área A se aplica una fuerza normal F de manera uniforme, la presión, p , viene dada por³²:

$$p = \frac{F}{A} \quad (1.1)$$

Sin embargo, el caso más general es aquel donde la fuerza puede tener cualquier dirección y que no se encuentre uniformemente distribuida en todos los puntos de la superficie. Para este caso más general la presión queda definida como:

$$p = \frac{d\mathbf{F}}{dA} \cdot \mathbf{n} \quad (1.2)$$

donde n representa un vector unitario normal a la superficie donde se pretende medir la presión. Queda claro de esta expresión que p es una magnitud escalar y por tanto no tiene dirección. Pese a que en el lenguaje cotidiano nos acostumbremos a hablar de que la presión se ejerce en una dirección u otra, es la fuerza que actúa la que tiene direccionalidad mientras que la presión que se genera en la superficie que soporta dicha fuerza resulta ser adireccional. La ecuación (1.2) nos muestra una relación inversamente proporcional entre la presión y la superficie donde esta se aplica, lo que demuestra complejidad intrínseca de realizar experimentos a presiones elevadas en muestras de gran tamaño.

La presión según el sistema de unidades internacionales (SI) se mide en Pascales (Pa); un Pascal equivale a una fuerza de 1 Newton actuando sobre un metro cuadrado. Sin embargo, la presión es posiblemente la magnitud física que puede ser medida en un mayor número de unidades diferentes. En la Tabla 1.1 mostramos diferentes tipos de unidades habituales de medida de la presión y las relaciones existentes entre ellas.

1.2.1. Presión hidrostática

El hecho de que la presión sea la medida de una fuerza por unidad de área hace que existan multitud de presiones diferentes en función del campo en el que nos desen-

	Pa	bar	atm	Torr
1 Pa (N/m^2)	1	10^{-5}	0.987×10^{-5}	0.0075
1 bar	10^5	1	0.987	750
1 atm	101325	1.01325	1	760
1 Torr ($mmHg$)	133.32	1.3332×10^{-3}	1.32×10^{-3}	1

Tabla 1.1: Diferentes unidades de presión y factores de conversión.

volvamos. Así, hablamos de presión atmosférica para referirnos a la presión que ejerce la atmósfera sobre la superficie terrestre; de presión de vapor, que indica la presión que ejerce el vapor de una sustancia sobre una fase condensada cuando ambas se encuentran en equilibrio termodinámico. Pero hablamos también de presión arterial, presión ocular, presión de un gas ideal o presión de un fluido en función del objeto que ejerce la fuerza y la superficie sobre la que se ejerce en cada momento.

Sin embargo, existe una clasificación general para todas ellas en función de la dirección en la que actúan las fuerzas que generan la presión. Hablaremos de **presión uniaxial** o estática si la fuerza que se ejerce sobre una determinada muestra le es transmitida en una dirección determinada o preferencial. En cambio hablamos de **presión isostática** cuando la fuerza a la que es sometida una muestra determinada es la misma en todas las direcciones del espacio. Específicamente, tratamos con una **presión hidrostática** cuando la presión isostática es transmitida a través del agua o cualquier otro fluido³³.

En este trabajo, cuando nos referiremos a presiones, nos referiremos siempre a presiones hidrostáticas, puesto que las proteínas en el interior de los seres vivos se encuentran rodeados de disolvente, principalmente agua, y por tanto cuando el disolvente ejerce una fuerza contra la proteína lo hace siempre dentro de un régimen isostático, donde la presión se está ejerciendo sobre todos los puntos de la superficie de la proteína. Cuando experimentalmente se trabaja con proteínas a alta presión, la presión medida también es referida a una presión hidrostática alrededor de toda la superficie de la proteína.

1.2.2. Aplicaciones de la presión como magnitud termodinámica

Es interesante señalar que la investigación con la presión es una ciencia relativamente joven. Tradicionalmente, la mayor parte de la producción científica, al menos desde un punto de vista bioquímico, ha estado ligada al uso de la temperatura como

variable termodinámica. No obstante, desde hace algunas décadas la presión ha adquirido un interés creciente, seguramente debido a una cuestión técnica, puesto que cada vez existe una mayor viabilidad de cara a poder contar con dispositivos con los que poder modificar la presión, y sobre todo, con los que poder medir la presión a la que se encuentra el sistema. Esto ha llevado a que cada vez sean más los trabajos y estudios relacionados con esta magnitud física³⁴, especialmente en el campo de las altas presiones, por encima de los 1000 bar, dando lugar a aplicaciones de distinta índole³⁴, en parte porque gracias a la presión se pueden reproducir de una manera mucho más sencilla, o más económica, ciertas propiedades similares a las que se obtienen mediante otras magnitudes termodinámicas como la temperatura. Además, se van descubriendo propiedades específicas de ciertos materiales que se generan al tratarlos con altas presiones, y que no existen cuando tratamos a estos materiales de una forma alternativa. Un buen ejemplo de esto último es la síntesis del diamante a alta presión, ya que un aumento considerable de temperatura y presión en el grafito nos conduce a un cambio de fase del carbono. Sea como fuere, la presión debe entenderse como una herramienta de gran valor, capaz de generar y proporcionar nuevas teorías y conocimientos que en muchos casos sentarán las bases de futuros descubrimientos, mientras que en otras situaciones servirán para comprender de una forma más completa y correcta algunos fenómenos ya descritos. Por esta razón, la investigación con presión no debe considerarse sólo como una ciencia aislada con aplicaciones directas, sino más bien como una ciencia interdisciplinar que permite llevar a cabo multitud de estudios incluidos en un amplio espectro científico. Por último, no conviene olvidar que la relación presión-temperatura está estrechamente ligada a través de la ecuación de estado de cualquier sustancia y por tanto existen muchos casos donde ambas magnitudes no son fácilmente separables, y trabajar a altas presiones implica en muchas ocasiones trabajar a altas temperaturas.

1.3. Influencia de la presión en el plegamiento de proteínas

El efecto que ejerce la presión sobre las proteínas es conocido desde principios del siglo XX, cuando Bridgman demostró con sus experimentos cómo era capaz de coagular la albúmina del huevo sin necesidad de un aporte calórico³⁵. La coagulación del huevo es un proceso de desnaturalización donde las proteínas de la clara del huevo, albúminas, pierden su estructura nativa para pasar a un estado desnaturalizado. En este proceso de desnaturalización existe un cambio de estado de la clara del huevo desde una fase líquida hasta una fase sólida, por agregación de las proteínas desnaturalizadas. De esta manera quedaba demostrado cómo un aumento en la presión del sistema produce resultados equivalentes al aumento de la temperatura. Sin embargo, la complejidad de obtener resultados experimentales a alta presión, especialmente en un ambiente biológico, ha llevado a conocer en menor medida el efecto de la presión con respecto a otros efectos que desnaturalizan las proteínas como temperatura, pH o agentes químicos. El continuo desarrollo de la biotecnología ha llevado a que en esta última década hayan sido varios los grupos de investigación que se han interesado por este tema por varios motivos^{36,37}. Por un lado existe la necesidad de seguir conociendo las fuerzas que gobiernan el plegamiento de una proteína sea cuales sean las condiciones en las que se encuentre el sistema^{38,39}. Pero además existen diferentes aplicaciones relacionadas con la presión y las proteínas que se han ido desarrollando en los últimos años. A continuación mostraremos algunos de los aspectos más importantes que se conocen a día de hoy sobre efecto que produce un aumento de la presión hidrostática en un medio en el que encontramos proteínas en disolución.

1.3.1. El diagrama de fases p - T de las proteínas

En términos generales, un diagrama de fases consiste en la representación gráfica de las fronteras entre diferentes estados de la materia de un sistema, en función de variables elegidas para facilitar el estudio del mismo. Por esta razón puede parecer extraño hablar de un diagrama de fases de las proteínas, cuando habitualmente no sufren cambios en su estado de la materia dentro de las condiciones fisiológicas. Sin embargo, si consideramos la proteína como un sistema que se puede encontrar en dos estados diferentes, el estado nativo, que podríamos considerar como un estado ordenado y en cierta forma ‘sólido’, y el estado desnaturalizado, considerado como un estado fundido y por tanto ‘líquido’, la representación de la estabilidad de dichos estados en función de variables seleccionadas se puede llegar a considerar como el diagrama de fases de las proteínas⁴⁰, donde cada línea muestra las fronteras entre el estado nativo y un estado desnaturalizado.

Desde finales del siglo XX es conocido el diagrama de fases p - T de las proteínas⁴¹. En la Figura 1.6 observamos una representación esquemática de este diagrama, donde se muestra la dependencia de la estabilidad de los estados nativo y desnaturalizado con respecto a la presión y temperatura. Se puede observar cómo el diagrama tiene forma elíptica. En el interior de la elipse encontramos el estado nativo y alrededor de esta elipse se encuentran los diferentes estados desnaturalizados. Antes de nada hay que recordar que el diagrama que se muestra es un diagrama esquemático, y que cada proteína mostrará un diagrama característico, donde la elipse podrá cambiar su tamaño, curvatura y orientación, manteniendo siempre una forma aproximadamente elíptica^{42,43}. Estudios realizados por diferentes grupos de investigación muestran que la mayoría de las proteínas estudiadas desnaturalizan dentro de un intervalo de entre 2 y 7 kbar de presión^{40,42–44}.

Lo primero que queda patente al analizar la imagen es que el estado biológicamente activo sólo es estable en unas condiciones determinadas de presión y temperatura, y por tanto el proceso de desnaturalización puede ser debido exclusivamente a la temperatura

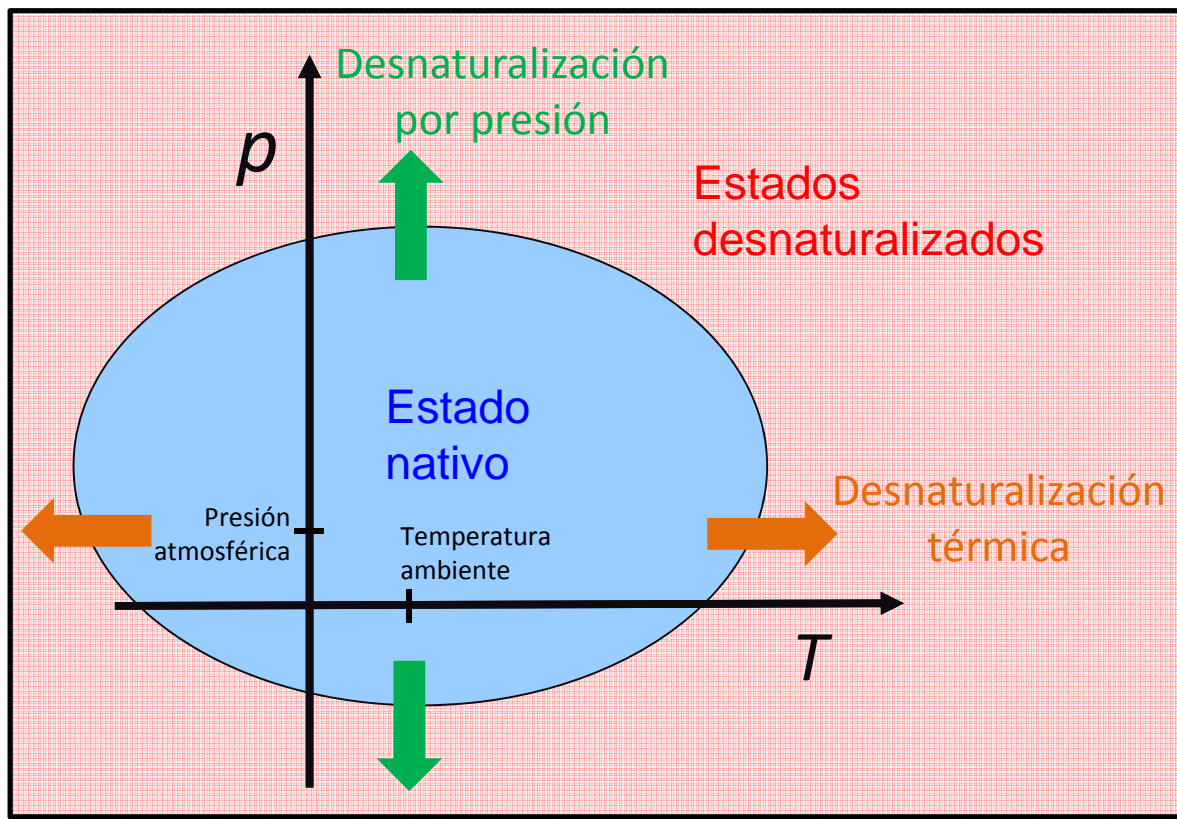


Figura 1.6: **Diagrama de fases p - T de las proteínas.** Se muestra en azul el estado nativo y en rojo los diferentes estados desnaturalizados. Las flechas naranjas muestran procesos de desnaturalización debidos a la temperatura a presión atmosférica, mientras que las flechas verdes muestran procesos de desnaturalización debidos a la presión a temperatura ambiente.

(flechas naranjas), debido exclusivamente a la presión (flechas verdes), o debido a una combinación de presión y temperatura cuando nos movemos en el resto de direcciones.

Nosotros enfocaremos nuestro trabajo a estudiar el plegamiento de proteínas en la zona superior derecha de este diagrama, la que coincide con presiones y temperaturas superiores a las condiciones ambientales. Podemos observar cómo para condiciones por encima de la temperatura ambiente y presión atmosférica, podemos obtener una desnaturalización debida a las altas temperaturas, **desnaturalización térmica**, o una desnaturalización debida a las altas presiones, **desnaturalización por presión**. Entre estos dos extremos también encontramos procesos de desnaturalización debido a la combinación de ambas variables. Lo más interesante de estos dos tipos de desnaturalización es que se producen de manera diferente, a través de mecanismos distintos, que

además nos llevan hasta estados desnaturalizados con diferentes propiedades⁴⁵. El estado desnaturalizado que se considera habitualmente se caracteriza por ser una cadena polipeptídica abierta sin apenas elementos de estructura secundaria formados. Este tipo de estado desnaturalizado es el que se describe para procesos de desnaturalización debidos a temperatura, pH o agentes químicos^{46,47}, y que vamos a denotar como **estado desnaturalizado ‘clásico’**. Corresponde al estado desnaturalizado que obtendríamos al salirnos del estado nativo en el diagrama de fases por el lado derecho de la elipse a través de un proceso de desnaturalización térmica. Pero estudios recientes⁴⁸ muestran que cuando salimos de la elipse por la parte superior, a temperatura ambiente y altas presiones, lo que obtenemos son estructuras bastante compactas, que mantienen cierta apariencia de estructura nativa, ya que siguen manteniendo una estructura globular. Son conformaciones con los elementos de estructura secundaria formados, pero con una mayor distancia entre ellos. De esta forma ocupan un mayor volumen global que el estado nativo, por lo que se trata de un **estado desnaturalizado ‘hinchado’**, dando la sensación de que fuese una estructura similar a la nativa pero expandida desde su interior^{39,49,50}. Este tipo de estados son característicos de la desnaturalización dirigida por la presión. Por tanto, parece evidente que las condiciones que conducen a la desnaturalización de las proteínas son las responsables de las características del estado final.

1.3.2. Diferentes mecanismos de desnaturalización

El mecanismo por el cual las proteínas se desnaturalizan por efecto de la temperatura es conocido desde hace muchos años^{51,52}. Lo podemos resumir de la siguiente manera: cuando existe un aumento de la energía del sistema, debido al aumento de la temperatura, las interacciones responsables de mantener la estructura terciaria de la proteína se debilitan notablemente; recordemos que, como hemos comentado anteriormente, estas fuerzas son mayoritariamente interacciones hidrófobas entre cadenas laterales, que

consideradas individualmente no son especialmente energéticas. Cuando la energía es suficientemente alta, y alcanzamos por tanto a la temperatura de transición, las interacciones que mantienen la estructura nativa son tan débiles que la proteína no es capaz de estabilizar este estado y la proteína pierde su estructura globular. En ese momento, además, se van a ir perdiendo parte de los enlaces de hidrógeno intramoleculares que mantenían los elementos de estructura secundaria. Aunque estos enlaces son más fuertes por tratarse de enlaces de hidrógeno y no de interacciones entre residuos hidrófobos, su formación estaba favorecida por el acercamiento de ciertos aminoácidos gracias al empaquetamiento de las cadenas laterales, y al perderse parte de estas interacciones la cadena polipeptídica no adopta la conformación idónea para que se establezcan estos enlaces entre grupos de la proteína. Además, en el momento que la proteína pierde su estructura globular, las moléculas de disolvente de carácter polar tienen cierta accesibilidad a los aminoácidos interiores que mantienen los elementos de estructura secundaria y pueden formar enlaces de hidrógeno con estos, compitiendo con los enlaces que mantienen la estructura secundaria. Cuando la temperatura es superior a la temperatura de transición, lo que observamos es la proteína como una cadena lineal de aminoácidos alcanzando aproximadamente la estructura conocida como *random coil*, que coincide con la visión ‘clásica’ de la proteína desnaturalizada.

Sin embargo, aunque a día de hoy no se conoce con exactitud el proceso por el cual un aumento de la presión nos lleva al estado desnaturalizado, se postula un mecanismo diferente al que acabamos de explicar⁵³. Según este mecanismo, el proceso de desnaturalización por presión se da de la siguiente manera: al aumentar la presión hidrostática del medio se genera una fuerza sobre el sistema, en nuestro caso la proteína. La presión ejerce una fuerza que empuja al disolvente hacia el interior de la proteína. Si bien hemos descrito la proteína plegada como una estructura compacta, donde el núcleo hidrófobo queda en el interior y expone hacia el exterior sus aminoácidos de carácter polar, esta estructura no es totalmente maciza, y por tanto a medida que aumenta la presión las

moléculas de agua que rodean la superficie de la proteína se van introduciendo en el interior de la estructura globular, ocupando pequeños huecos o **cavidades** vacías en el interior de la proteína. La ocupación de estas cavidades no tiene lugar a presión atmosférica, puesto que la hidrofobicidad del interior de la proteína impide la entrada de estas moléculas en su interior. Sin embargo, cuando la presión es suficientemente alta, la fuerza que empuja a las moléculas al interior se hace superior a la interacción hidrófoba que las mantiene en el exterior. Las moléculas de disolvente en el interior de la estructura proteica dan lugar a un debilitamiento progresivo de la estructura terciaria, ya que no sólo se pierde parte de las interacciones hidrófobas que mantienen la estructura nativa, sino que al mismo tiempo las moléculas de agua van generando enlaces de hidrógeno con algunos aminoácidos, debilitando a su vez los enlaces de hidrógeno intramoleculares. De esta manera, la estructura nativa de la proteína se va perdiendo para dar lugar a una estructura desnaturalizada, con diferentes propiedades a las de la estructura tipo *random coil*. Se trata de la que conocemos como estructura ‘hinchada’, típica de los procesos de desnaturalización por presión^{??}. Es similar a la estructura nativa, pero con un volumen interno ocupado por moléculas de agua. Cuando la presión del medio continúa aumentando, la estructura sigue aceptando moléculas de agua en su interior al mismo tiempo que continúa debilitando las interacciones que mantienen la estructura globular. Cuando la presión es muy alta, las interacciones nativas se encuentran completamente debilitadas y la proteína no es capaz de estabilizar su estructura globular, por tanto la proteína evoluciona desde el estado ‘hinchado’ hasta una conformación totalmente desestructurada tipo *random coil*. Por tanto, mediante este mecanismo queda explicada la existencia de una estructura desnaturalizada que difiere de la estructura desnaturalizada ‘clásica’, al menos dentro de un intervalo de presiones relativamente moderado.

1.3.3. Factores que afectan a la desnaturalización mediada por la presión

Una vez comprendido, al menos de forma global, el mecanismo por el cual una proteína se desnaturaliza debido al efecto de la presión, se han intentado estudiar por varios métodos las diferentes propiedades que se ven alteradas a nivel local durante esta transición⁵⁴. Desde un punto de vista termodinámico, ya desde finales del siglo XX se conoce la principal propiedad de las estructuras desnaturalizadas por presión. Los estados desnaturalizados por presión (considerando conjuntamente la proteína junto con las moléculas de disolvente) tienen un volumen molar parcial inferior al de la misma proteína en su estado nativo^{55,56}. Esto, unido al hecho de que según el principio de Le Châtelier un incremento de la presión en el sistema nos lleva a un desplazamiento en el equilibrio hacia la especie de menor volumen⁵⁷, explica el hecho de que un aumento de la presión nos lleve hasta la estabilización del estado ‘hinchado’. La disminución del volumen molar parcial del estado desnaturalizado por presión se puede explicar en base a una serie de circunstancias que se dan a medida que aumenta la presión hidrostática del medio⁵⁸.

La estabilidad de un estado frente a otro va a depender de dos factores, la estabilidad entrópica y la estabilidad entálpica. Como comentamos anteriormente, el efecto hidrófobo es la fuerza principal a la hora de dirigir el plegamiento hacia la estructura nativa. La interacción entre residuos hidrófobos reduce notablemente la superficie de grupos apolares en contacto directo con el disolvente, de naturaleza polar, lo que estabiliza entrópicamente al estado nativo con respecto al estado desnaturalizado. El efecto hidrófobo produce una ordenación local de las moléculas de agua que se sitúan alrededor de la cadena polipeptídica, puesto que las moléculas de agua en contacto con los residuos hidrófobos tienen un índice de coordinación inferior y se encuentran más ordenadas que las moléculas que se encuentran alejadas de los centros hidrófobos. La diferencia de densidad local existente entre las moléculas de agua cercanas a los residuos

hidrófobos y las moléculas de agua que se encuentran embebidas en el disolvente es la que, a presión atmosférica, genera estabilidad del estado nativo con respecto al estado desnaturalizado⁵⁹. Sin embargo, a medida que aumenta la presión, los espacios internos de la proteína se van rellenando de moléculas de agua dando lugar a que el empaquetamiento interno de la proteína se haga mayor, cuando consideramos conjuntamente la proteína y el disolvente. Cuando el empaquetamiento del interior de la proteína es suficientemente elevado, debido a que las moléculas han ido accediendo desde el exterior hacia el interior de la proteína, existe una homogeneización de la densidad entre las diferentes moléculas de disolvente⁶⁰, y esto lleva a que la diferencia de estabilidad entre el estado nativo y el estado desnaturalizado bajo presión sea muy inferior a la que existía entre el estado nativo y el estado desnaturalizado a presión atmosférica. La desestabilización del estado nativo con respecto al estado hinchado nos lleva a que cuando la presión es suficientemente alta, el estado hinchado pase a ser el más estable.

Pero además del factor entrópico que favorece la desestabilización del estado nativo con respecto al estado desnaturalizado, recientemente se ha estudiado otro factor importante que influye de forma considerable en el proceso de desnaturalización debido a la presión: **las cavidades en el interior de las proteínas plegadas**^{61,62}. Las cavidades de las proteínas globulares son intrínsecas a su estructura, ya que el propio empaquetamiento de la cadena principal en el plegamiento de la proteína da lugar a pequeños volúmenes vacíos en el interior de su estado nativo. El número, la forma, la situación y el tamaño de las cavidades internas de la proteína afecta considerablemente al proceso de desnaturalización, puesto que proteínas con cavidades mayores en su interior se ven más afectadas por el efecto de la presión que proteínas más compactas en su interior⁵⁸. Cuando se encuentra bajo una presión hidrostática, la proteína va a sufrir una fuerza neta sobre su superficie. Las cavidades que tenga la proteína en su interior van a ser zonas internas de la proteína donde hay una menor densidad al ser menos compactas, y por tanto van a ejercer una menor resistencia frente a la presión si la comparamos con

otras proteínas con cavidades de menor tamaño y por tanto más densas en su interior. Por tanto existe una relación entre cuan compacta es una estructura y cuánto afecta la presión hidrostática a su estabilidad.

De esta manera, hemos hecho un pequeño repaso a los antecedentes que tenemos a día de hoy acerca del plegamiento de proteínas bajo una presión hidrostática creciente, dando especial relevancia a los factores termodinámicos.

1.3.4. Aplicaciones de la desnaturalización mediada por presión

Una aplicación especialmente interesante de la presión es **impedir la formación de agregados proteicos**^{63,64}. Los agregados proteicos, entre los que se encuentran los agregados β -amiloides, se producen por un mal plegamiento de las proteínas debido a una alta concentración de proteína. Dichos agregados están formados por hebras β que se asocian entre sí, de una manera similar a como lo hacen las láminas β en el estado nativo de la proteína. Los agregados pueden crecer bastante hasta formar fibras amiloides insolubles, las cuales interfieren en el correcto funcionamiento de la célula dando lugar a lo que se conoce como las enfermedades del plegamiento^{65,66}. Este tipo de enfermedades, algunas de ellas neurodegenerativas entre las que se encuentran el Alzheimer, Parkinson o Huntington^{67,68}, están cada vez más presentes en la sociedad actual, al ser una sociedad de mayor edad y más protegida contra otro tipo de enfermedades. Son enfermedades que despiertan un gran interés puesto que a día de hoy sus causas son bastante desconocidas y carecen de ningún tipo de tratamiento definitivo. Desde que quedó demostrado que a través de un aumento de la presión hidrostática en el medio, los primeros agregados proteicos formados se pueden disociar recuperando el estado nativo de la proteína^{63,69}, la presión ha supuesto una herramienta esencial para comprender completamente el mecanismo por el cual se comienzan a producir estas enfermedades debidas al mal plegamiento. Además de las posibles aplicaciones en medicina, la utilización de la presión para disociar los agregados proteicos presenta ciertas aplicaciones en

otros campos de la ciencia⁷⁰. **En farmacología**⁷¹, por ejemplo, es cada vez más común el uso de péptidos o fragmentos de proteínas como medicamentos, debido a su alta eficiencia⁷². Sin embargo, la desactivación de estos medicamentos debido principalmente a la formación de agregados es su mayor inconveniente de cara a su fabricación industrial. La utilización de la presión como herramienta para rescatar el estado biológicamente activo, y por tanto eficaz del medicamento, resulta ser un buen método para la industria farmacéutica de cara a preservar el medicamento en las mejores condiciones⁷³.

Fuera del uso de la presión frente a la formación de agregados encontramos un uso creciente de la presión en **la industria alimentaria**^{74,75}. La presión puede ser usada como herramienta tanto para la conservación como para el tratamiento de muchos alimentos⁷⁶. Otra gran aplicación del plegamiento de proteínas modulado por presión la encontramos en la investigación básica, especialmente **en el estudio del embudo de plegamiento**, ya que la transición entre el estado nativo y el estado desnaturalizado por presión resulta ser normalmente mucho más rápido que el existente entre el estado nativo y el estado desnaturalizado térmicamente⁷⁷. Por tanto el uso de la presión resulta una herramienta muy adecuada para poder estudiar esta transición, ya que reduce considerablemente los tiempos de plegamiento de la proteína, facilitando el estudio cinético del proceso⁷⁸.

De esta manera hemos hecho un repaso de los antecedentes que encontramos para sistema proteína-presión. Estos deben de ser el punto de partida sobre el que se base este trabajo para seguir complementando el conocimiento acerca de este sistema.

1.4. Estudio del plegamiento de proteínas

Para intentar responder a las preguntas que hay todavía sin respuesta el problema del plegamiento, se ha estudiado este tránsito por dos vías principales: a través de los insustituibles experimentos y a través de aproximaciones teóricas. Dentro del campo experimental, la aproximación al plegamiento se ha realizado a su vez por dos métodos

diferentes. Por un lado existe el estudio utilizando técnicas de tipo biofísico, mediante las cuales se monitoriza alguna propiedad del sistema mientras se pliega y desnaturaliza la proteína⁷. Una de las técnicas más modernas pero con grandes perspectivas dentro de este campo son los experimentos de **molécula única**⁷⁹. Estas técnicas permiten considerar una única molécula y realizar su seguimiento en el tiempo mientras se mide algún tipo de interacción con el medio, por ejemplo el proceso de plegamiento de una proteína⁸⁰. La segunda aproximación al plegamiento desde un punto de vista experimental es la denominada **ingeniería de proteínas**⁸¹, mediante la cual en el laboratorio se pueden sintetizar secuencias de aminoácidos específicas, entre las que predominan la generación de secuencias mutadas^{7,82}, con las que se pueden estudiar las consecuencias cinéticas y termodinámicas del cambio de secuencia en el proceso de plegamiento.

Por otro lado, existe una aproximación teórico-computacional al plegamiento de proteínas que se ha desarrollado en los últimas décadas, y a su vez se divide en dos vías principales en función del problema que se quiere estudiar. Una forma de abordar el plegamiento es desde un punto de vista predictivo. Existe una parte de la comunidad científica que utiliza la modelización molecular para intentar **predecir la estructura nativa de la proteína** a partir de su secuencia de aminoácidos⁸³. La predicción del estado nativo de la proteína es un problema que sin embargo sigue sin estar resuelto⁸⁴ a pesar de los numerosos avances en este campo y del increíble avance que supondría dentro del estudio de las proteínas una herramienta que permitiese conocer la estructura tridimensional nativa de la proteína únicamente a partir de su secuencia. Un segundo punto de vista para estudiar el plegamiento a través de métodos teórico-computacionales⁸⁵ es **el estudio del proceso de plegamiento en sí mismo**. Este enfoque del estudio del plegamiento consiste no tanto en predecir la estructura tridimensional de la proteína sino más bien en un estudio completo del mecanismo por el que pliega la proteína, siguiendo y estudiando las características de la evolución de esta desde su estado desnaturalizado hasta su estado nativo. De esta manera se puede estudiar la termodinámica y la cinética

del proceso, así como la posible aparición de estados intermedios durante el proceso de plegamiento.

Desde una aproximación experimental, el trabajo con proteínas a altas presiones se encuentra todavía en una fase muy inicial debido a los problemas de tipo técnico que supone la utilización de las técnicas de alta presión con la materia blanda. Un ejemplo lo podemos encontrar en la obtención de información estructural para proteínas desnaturalizadas por presión. Existen dos técnicas principales para la obtención de la estructura de proteínas, la cristalografía de rayos X y la Resonancia Magnética Nuclear (RMN). Ambas técnicas son las que se utilizan también para el estudio de la estructura a alta presión.⁸⁶ Para poder realizar un estudio por cristalografía de rayos X es necesario obtener un cristal de la estructura a resolver⁸⁷. La obtención de este cristal ya resulta bastante complicada en ciertas ocasiones a presión atmosférica, y su obtención a presiones elevadas es si cabe más compleja. En los últimos años se han introducido algunos métodos de ‘crío-refrigeración’, mediante los cuales es más sencillo la obtención de la estructura cristalina^{88,89}. Por tanto, aunque existen algunas estructuras cristalizadas a alta presión⁹⁰, estas son las menos en comparación con las estructuras cristalinas a presión atmosférica. Por tanto, para un estudio de una estructura determinada a altas presiones, parece más adecuada la utilización de RMN⁹¹, que pese a presentar ciertos inconvenientes a la hora de interpretar los desplazamientos químicos a alta presión, a día de hoy resulta ser la mejor técnica para elucidar la conformación de una proteína sometida a alta presión. Por tanto existe un mayor número de estructuras de proteína a alta presión obtenidas por esta técnica que por la de rayos X^{92,93}. Además al ser RMN una técnica dinámica, al contrario que los rayos X, que son una técnica estática, RMN permite un estudio dinámico de la evolución de la proteína a medida que aumentamos la presión⁹⁴, lo que permite hacer un estudio cinético del proceso de plegamiento, a la vez que permite reconocer intermedios cinéticos y termodinámicos, gracias a que la muestra puede ser alterada a la vez que se realiza el espectro. Dicho estudio no puede ser

realizado por rayos X, al necesitar la estructura cristalina de la proteína que se quiere estudiar en cada momento.

Por estos motivos, el plegamiento bajo presión hidrostática ha sido mucho más estudiado a través de la simulación molecular, que permite eliminar las dificultades técnicas que supone el uso de altas presión con la materia blanda. Además, el uso de la simulación permite no sólo reproducir a través de un ordenador el sistema proteína-presión y el proceso de plegamiento de forma dinámica⁹⁵⁻⁹⁷, sino que a través de la simulación se han podido ir estudiando diferentes factores, que han sido explicados a lo largo de esta Introducción y que afectan de forma decisiva a la desnaturalización de proteínas a altas presiones. Por ejemplo, se ha estudiado la disminución del volumen molar parcial del estado desnaturalizado al aumentar la presión⁹⁸, el estudio de cavidades en el interior de una proteína⁹⁹ o el mecanismo por el cual la inyección de moléculas de agua en el interior de la proteína nos lleva hasta el estado desnaturalizado¹⁰⁰. De esta forma, la simulación se ha convertido la mayoría de las veces en el mejor método de obtener información de estos sistemas, puesto que muchos de estos procesos son difícilmente medibles en un laboratorio, al menos con las técnicas que existen hoy en día.

Nosotros abordaremos el estudio del proceso de plegamiento a través de la simulación molecular. Estudiaremos la termodinámica de este proceso, dejando de lado el punto de vista predictivo. Pretendemos desarrollar un modelo que nos permita estudiar la termodinámica de la desnaturalización inducida por presión. Para ello, es necesario tener un modelo que describa la proteína y sus interacciones, y un algoritmo de muestreo conformacional eficaz¹⁰¹. A lo largo del próximo Capítulo describiremos con detalle el modelo que hemos desarrollado para poder reproducir el efecto de la presión en el plegamiento de proteínas.

1.5. Objetivos y organización de la Tesis

El objetivo principal de esta Tesis es la generación de un modelo de simulación molecular que nos permita reproducir y estudiar el efecto que supone un aumento de la presión del sistema durante la transición de plegamiento/desnaturalización de las proteínas globulares. El modelo no sólo será desarrollado desde un punto de vista teórico, sino que además será probado y optimizado con la simulación de una proteína ampliamente estudiada y conocida en nuestro grupo de investigación. Una vez hayamos desarrollado un modelo que nos permita estudiar al sistema a diferentes presiones y temperaturas, realizaremos simulaciones de diferentes proteínas buscando similitudes y diferencias entre los diferentes caminos de plegamiento que vayamos obteniendo. Daremos especial importancia a dos factores principales: por un lado al estudio a nivel local de los diferentes estados desnaturalizados obtenidos y por otro lado al efecto que produce un aumento de presión en los dos elementos de estructura secundaria más frecuentes en las proteínas, las láminas β y las hélices α . Por último, planteamos la simulación de una proteína cuyo comportamiento a altas presiones ha sido estudiado tanto de forma teórica como de forma experimental. Nuestra intención es poder comprobar con resultados experimentales si los resultados obtenidos por nuestro modelo reproducen los efectos reales que se observan en la proteína al someterla a alta presión. De esta manera podremos confirmar si nuestro objetivo principal ha sido o no resuelto de forma satisfactoria.

La memoria de esta Tesis se ha dividido en cinco capítulos, además de las conclusiones, para que de una forma ordenada podamos ir respondiendo a todas los objetivos que acabamos de plantear.

A lo largo del **Capítulo 2**, nos introduciremos en el mundo de la simulación molecular, explicando de forma detallada el modelo que hemos diseñado. Dicho modelo se basará en los potenciales de tipo $G\bar{o}$ que nuestro grupo de investigación ha venido utilizando para estudiar el plegamiento a presión atmosférica en los últimos años¹⁰². A lo largo del capítulo explicaremos el proceso de muestreo de la superficie del embudo

de plegamiento, detallaremos el modelo geométrico y sobre todo el potencial de interacción que hemos utilizado. Además, explicaremos algunas de las herramientas que hemos utilizado para el análisis de los resultados obtenidos.

Una vez introducido de forma teórica el modelo que hemos generado, a lo largo del **Capítulo 3** mostraremos los resultados que hemos obtenido al aplicar dicho modelo sobre una proteína real. Este será el capítulo principal de esta Tesis, puesto que no sólo nos servirá para entender qué le sucede a la proteína estudiada cuando alteramos la presión del medio, sino que los resultados que mostremos a lo largo de este Capítulo servirán como calibrado de nuestro potencial. Al final de este Capítulo se realizará también un estudio estructural del estado desnaturalizado por presión, para así entender qué le sucede a la proteína a nivel local durante el proceso de desnaturalización. Así, una vez terminado ese Capítulo, comprenderemos cómo actúa nuestro modelo y hasta dónde puede llegar en el estudio del plegamiento.

El **Capítulo 4** está dedicado a estudiar una batería de proteínas con distintos tipos de estructura secundaria, hélice α o lámina β . De esta manera podremos entender cómo el modelo actúa sobre los dos principales elementos de estructura secundaria, al mismo tiempo que seguimos completando el número de sistemas estudiados, buscando características comunes y diferencias entre los procesos de plegamiento estudiados para diferentes proteínas.

Por último, a lo largo del último Capítulo de resultados, **Capítulo 5**, realizaremos un estudio del plegamiento de una proteína muy estudiada tanto desde el punto de vista de la simulación como a través de experimentos. Así podremos comparar nuestros resultados con los de otros grupos de investigación, y de este modo poder analizar de manera completa y más cuantitativa el potencial que hemos diseñado en esta Tesis.

De esta manera habremos abarcado un estudio completo del efecto de la presión en diferentes proteínas y continuaremos ampliando las bases del conocimiento del sistema proteína-presión.

Capítulo 2

Materiales y métodos

El continuo desarrollo que han sufrido la informática en las últimas décadas ha llevado a que cada vez sea más asequible adquirir ordenadores de gran capacidad. Esto nos da la posibilidad de desarrollar cada vez más una vía de investigación, **la simulación**, mediante la cual podemos reproducir experimentos reales o predecir fenómenos de la naturaleza a partir de las ecuaciones que describen la física y la química de los sistemas que deseamos estudiar. La simulación se convierte en un tercer pilar fundamental de un investigador, junto a la teoría y a los experimentos, y uniendo estos tres conceptos podremos obtener una información mucho más completa del sistema estudiado.

La simulación nos ofrece además un abanico de formas alternativas para el estudio de un sistema. Podemos aplicar técnicas de distinta índole que difícilmente se pueden encontrar en un mismo laboratorio. Por ejemplo, se puede simular un espectro de RMN en el mismo ordenador donde simulamos un espectro Raman, mientras que en un laboratorio se necesitarán dos equipos totalmente distintos para obtener estos dos espectros.

Pero la simulación también nos permite abordar el estudio de sistemas que son muy difíciles o imposibles de estudiar a través de experimentos en vivo, por ejemplo haciendo una simulación del interior de la Tierra, de la superficie de un planeta que se encuentra a años luz del sol o la simulación de ciertos experimentos que conllevan

el uso de productos nocivos para el ser humano, por ejemplo el uso de disolventes tóxicos o de productos radiactivos. Sin olvidarnos del factor económico, puesto que la experimentación es realmente cara, y los ordenadores son cada vez más baratos y de mayor capacidad. Por tanto parece interesante, si es posible, reducir el número de experimentos para estudiar un sistema a cambio de realizar simulaciones previas.

Cabe destacar que, aunque el coste económico de las máquinas para realizar las simulaciones moleculares se ha reducido drásticamente en los últimos años, en la simulación existe un coste en tiempo al que denominamos **coste computacional**. Se refiere al tiempo que tarda un ordenador en procesar los cálculos sobre el programa de simulación. El coste computacional será siempre un factor limitante. Si bien un mejor procesador implica un menor coste computacional, cada vez los programas de simulación son más complejos y precisos, y esto conlleva un mayor coste computacional. Por tanto se hace necesario encontrar un equilibrio entre la exactitud de los resultados y el coste computacional que eso conlleva, ya que no parece conveniente realizar de forma rutinaria simulaciones muy precisas pero que puedan llegar a tardar meses o años en completarse.

Por todo lo comentado, parece que la simulación molecular puede ser una herramienta cada vez más utilizada en ciencia, y por tanto la generación de modelos adecuados al problema a estudiar supone un nuevo desafío para muchos investigadores, ya que el verdadero reto para alguien que realiza una simulación consiste en **escoger el modelo más adecuado posible para estudiar un problema determinado**.

En este Capítulo se mostrarán las características del modelo que nosotros hemos desarrollado, así como las optimizaciones que hemos realizado en su implementación. En último lugar también mostraremos algunas de las herramientas de cálculo que hemos preparado para el análisis de los resultados obtenidos de nuestras simulaciones.

2.1. Motivación del Capítulo

Cuando estudiamos el proceso de plegamiento de una proteína, lo que queremos obtener es un conjunto de conformaciones que la proteína adopta cuando pasa de un estado desnaturalizado a un estado nativo. Cada una de estas conformaciones, como ya hemos visto en la Sección 1.1.2, suponen un punto en la superficie de energía del embudo de plegamiento. Cuando aplicamos una presión hidrostática al medio, lo que hacemos es forzar un cambio en esa superficie del embudo de plegamiento, de tal forma que el nuevo camino de plegamiento que adoptará la proteína supondrá una nueva ruta con distintas configuraciones asociadas. Esto queda reflejado en el esquema de la Figura 2.1. El poder muestrear esta nueva superficie del embudo de la forma mas representativa posible será absolutamente prioritario para así poder entender la termodinámica de este proceso.

En simulación se plantean dos vías principales para poder hacer un buen muestreo, **la Dinámica Molecular y el Método de Monte Carlo**^{103,104}. En la Dinámica Molecular, lo que se hace en primer lugar es definir un potencial de interacción entre todos los átomos o centros de interacción seleccionados. Cada una de estas interacciones refleja una de las distintas fuerzas que aparecen entre los diferentes centros de interacción. La integración numérica de la segunda Ley de Newton nos lleva a obtener las velocidades y posiciones de cada partícula en función de intervalos discretos de tiempo. Esto nos permite tener un ‘vídeo’ de nuestro sistema donde fotograma a fotograma podemos ir observando lo que le está sucediendo.

Sin embargo, cuando estudiamos fenómenos que involucran procesos muy rápidos en el tiempo, como es el caso del movimiento de átomos enlazados covalentemente¹⁰⁵, se hace necesario elegir una unidad de tiempo suficientemente pequeña como para poder seguir temporalmente a las partículas del sistema. En el caso de muestrear estructuras moleculares este intervalo puede ser del orden de 10^{-15} s. Esto nos lleva a la generación de un número enorme de configuraciones, especialmente cuando estudiamos un fenómeno bastante largo en el tiempo como es el plegamiento de proteínas¹⁰⁶. En la Figura 2.2

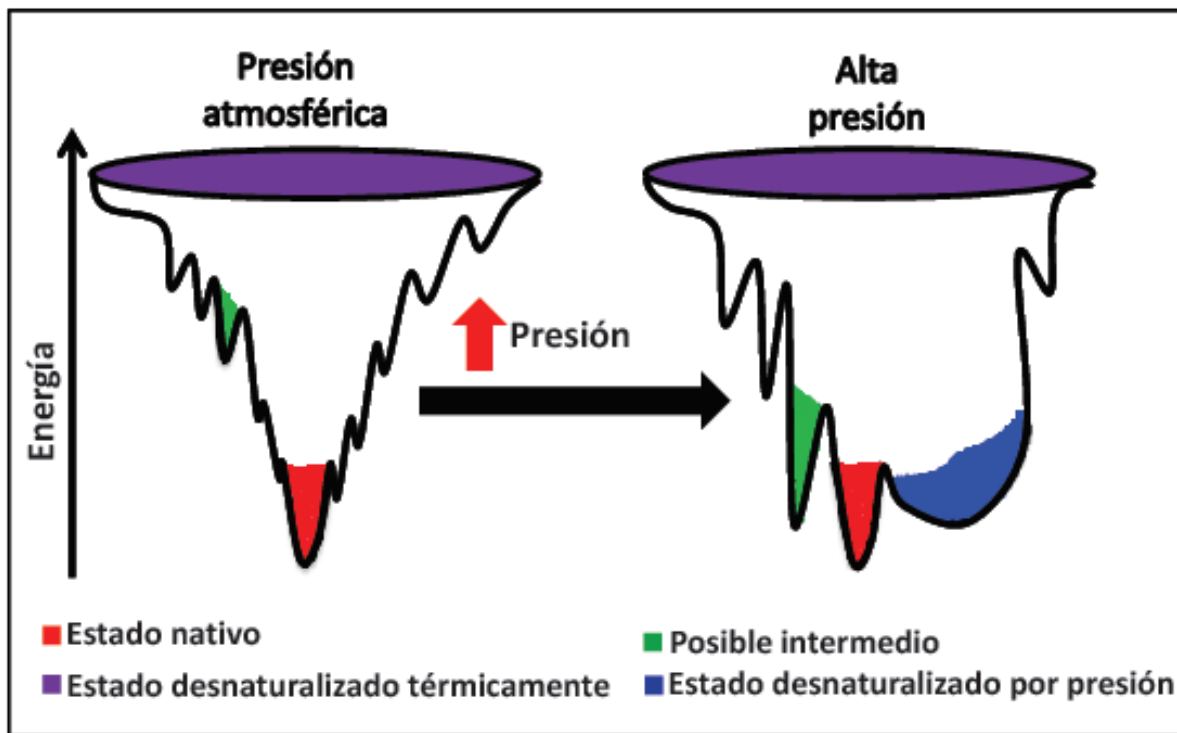


Figura 2.1: **Esquema que ilustra el objetivo principal de nuestra investigación.** Aplicamos un aumento de presión en el sistema para así modificar el embudo de plegamiento y poder estudiar las nuevas transiciones entre el estado nativo y el estado desnaturalizado.

podemos observar una escala de tiempos para los movimientos típicos de una proteína: las flechas en color rojo de la parte inferior indican tiempos reales para la simulación de estos movimientos mediante dinámica molecular, para una proteína de tamaño medio, en un medio acuoso (≈ 45000 moléculas de agua) realizados en un ordenador de escritorio (2.6 GHz dual-core).

Una solución para llevar a cabo un muestreo conformacional eficiente de un proceso muy largo en el tiempo es el Método de Monte Carlo, que resulta especialmente adecuado para estudiar procesos de equilibrio lentos. Este método consiste en generar configuraciones aleatorias dentro del posible espacio conformacional del sistema. Se utiliza un algoritmo de muestreo que genera configuraciones adecuadas y proporcionales en población a las que observaríamos a lo largo de la superficie del embudo de plegamiento. Con este método en principio se pierde toda información acerca de las velocidades de las partículas. Sin embargo, para un estudio termodinámico y estructural de ese proceso,

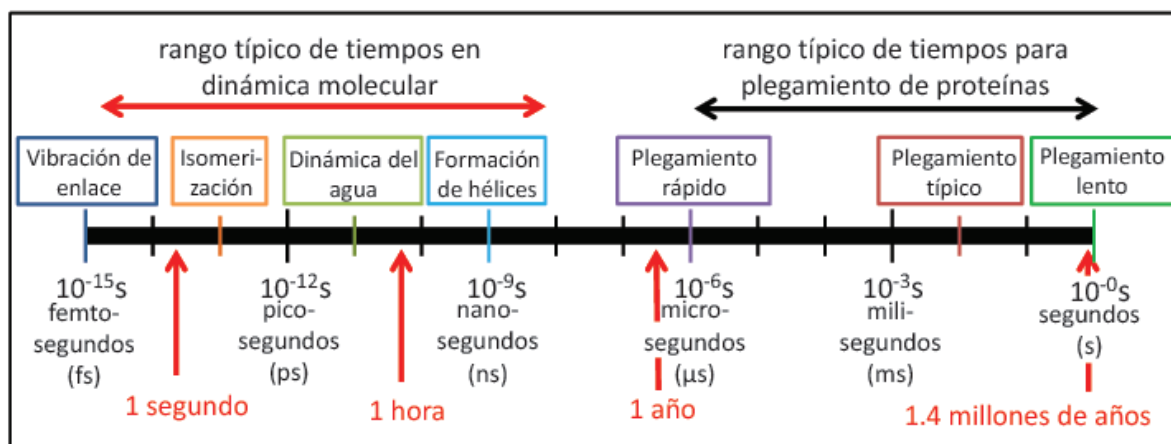


Figura 2.2: **Escala de tiempos de los movimientos típicos de una proteína.** Por encima de la escala de tiempos se muestran diferentes movimientos que se producen en el interior de las proteínas y el método de simulación que se suele emplear para la reproducción de estos movimientos. Por debajo de la escala de tiempos se muestra con una flecha roja el tiempo real, que tarda un ordenador de mesa habitual (2.6 GHz dual-core) en hacer estas simulaciones por Dinámica Molecular, para una proteína de tamaño medio con aproximadamente 45000 moléculas de agua como disolvente. Imagen adaptada de ^{105,107}.

esa falta de información no parece especialmente relevante. Al mismo tiempo, el número de configuraciones necesarias para muestrear el sistema completo no depende del tiempo total del proceso sino de la complejidad de la superficie de energía, necesitando una mayor cantidad de configuraciones cuando la superficie de energía es más compleja y por tanto existe una mayor variedad de configuraciones intermedias entre el estado nativo y el estado desnaturalizado.

Un buen ejemplo que muestra la diferencia entre estos dos métodos de simulación se puede plantear con el siguiente ejemplo. Supongamos que quisiéramos estudiar un sistema muy grande y desconocido para nosotros, por ejemplo el comportamiento de las hormigas en un hormiguero a lo largo de toda su vida. Podríamos hacerlo de dos formas distintas, podríamos monitorizar la vida completa de una única hormiga desde que nace hasta que muere, obteniendo información de lo que hace en cada momento a lo largo de su vida, lo que se correspondería a la Dinámica Molecular. O, por lo contrario, podríamos en un instante dado hacer una fotografía a cada uno de los individuos en el interior del hormiguero, de tal forma que obtendríamos imágenes de distintas hormigas en distintas etapas de sus vidas. Si el número de fotografías es suficiente, el análisis posterior de

esas imágenes nos lleva a resultados casi idénticos de los que habríamos obtenido por el primer método. Esta segunda manera de estudiar al sistema corresponde al Método de Monte Carlo.

En el estudio del plegamiento de proteínas, para que los resultados de la Dinámica Molecular fueran similares a los obtenidos por el método de Monte Carlo se haría necesario realizar un promedio sobre diferentes trayectorias en el tiempo. Esto conllevaría un elevado coste computacional y, a la postre, resulta ser el motivo principal por el que en este trabajo decidimos estudiar este proceso termodinámico y estructural a través del método de Monte Carlo. Además, aunque existen potenciales de interacción incluidos dentro de paquetes de simulación comerciales tanto para la Dinámica Molecular como para el método de Monte Carlo¹⁰⁸, que además nos podrían facilitar algunas fases del trabajo, preferimos desarrollar un potencial de interacción propio y ‘artesanal’, que nos permita abordar el estudio desde el punto de vista que deseemos, ya que un potencial desarrollado por nuestro grupo de investigación nos permite elegir en todo momento de qué manera queremos realizar el estudio de nuestro sistema.

Al mismo tiempo, para poder llevar a cabo nuestra simulación en un tiempo razonable, ya que las proteínas son macromoléculas de un tamaño considerable, lo más adecuado en nuestro caso es utilizar una **representación reducida** del sistema o **modelo *Coarse Grained***^{109,110}. Esto consiste en simplificar el modelo, reduciendo los grados de libertad del sistema que se desea estudiar. De esta manera se reduce notablemente el coste computacional de la simulación. Utilizando un modelo simplificado frente a otros más detallados perdemos cierta información, y con ello parte de resolución. Sin embargo, de cara a obtener resultados de un proceso tan completo y largo en el tiempo como el que estudiamos, es más que aceptable esta simplificación, siempre y cuando seamos conscientes de que los resultados no siempre podrán ser comparados con resultados experimentales de forma cuantitativa, sino más bien desde un punto de vista semi-cuantitativo o cualitativo. Esta última simplificación nos lleva inexorablemente a

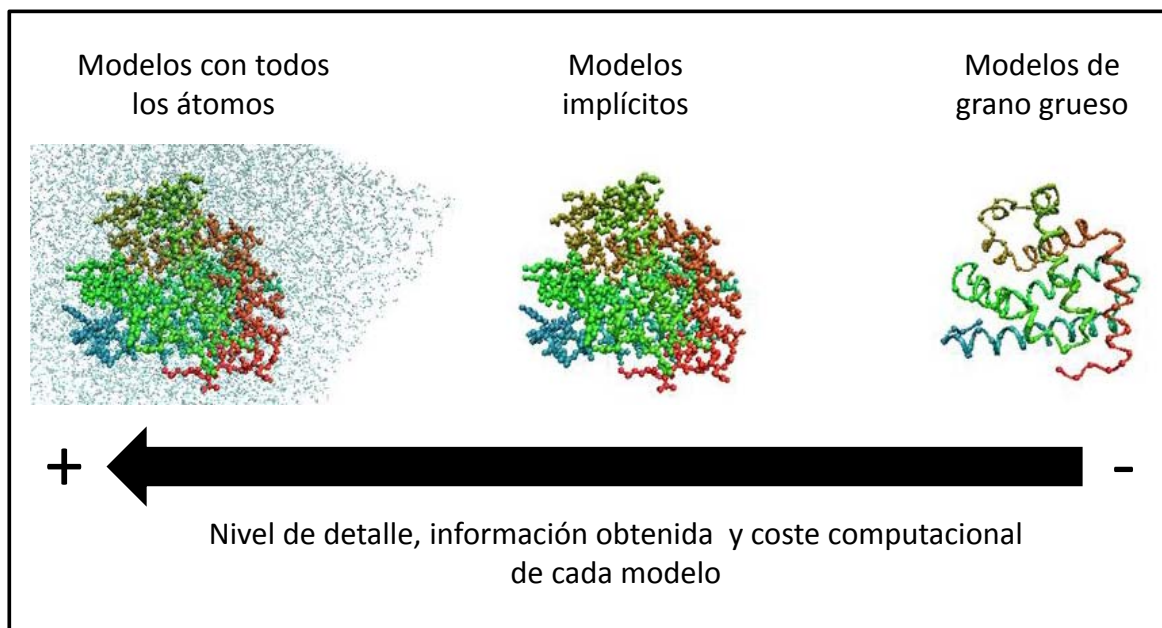


Figura 2.3: **Esquema de diferentes modelos con distinto nivel de detalle.** Se presentan tres modelos diferentes con distinta resolución para la misma proteína. Se observa cómo a medida que el modelo se simplifica, la resolución de la proteína es muy inferior. Imagen adaptada de¹¹¹.

tener que seleccionar potenciales de interacción adaptados al modelo simplificado que vayamos a utilizar. La Figura 2.3 muestra un esquema de distintos modelos con diferentes niveles de simplificación¹¹¹.

Por tanto, para llevar a cabo un estudio de la transición del plegamiento de proteínas mediante simulación molecular, es necesario diseñar un modelo de simulación que compatibilice la geometría del modelo con la resolución del potencial de interacción elegido, adecuándolo a un correcto método de simulación. A lo largo del Capítulo mostraremos de qué manera hemos podido conseguir todo este objetivo.

2.2. Modelo geométrico

Cuando utilizamos un modelo reducido para estudiar un sistema complejo, como es nuestro caso, debemos tener en cuenta dos factores fundamentales. Por un lado, por qué y para qué estamos utilizando un sistema reducido y, en segundo lugar, qué información hay que introducir en nuestro sistema simplificado para que este reproduzca de la forma

más precisa posible el sistema real, compensando en la medida de lo posible la reducción en el número de grados de libertad considerados. En nuestro caso utilizamos un modelo con **un único centro de interacción por residuo**, es decir, cada aminoácido se representa por un único centro, que se localiza en el carbono α ($C\alpha$) de cada aminoácido de la proteína. La distancia entre $C\alpha$ consecutivos en la secuencia es fija y de 3.8 Å, esta distancia corresponde a la distancia real que existe entre los $C\alpha$ de dos aminoácidos enlazados covalentemente. En la Figura 2.4 se puede observar cómo, para un fragmento de una proteína, con la estructura de $C\alpha$ reducimos drásticamente el número de centros de interacción si lo comparamos con la estructura atómica. De esta manera el número de grados de libertad del sistema se reduce sustancialmente. Recordemos que el número de átomos por aminoácido es bastante variable y va desde los 7 átomos en el caso de la glicina hasta los 24 átomos en el caso del triptófano, siendo el valor medio de aproximadamente 17 átomos por aminoácido. Por tanto, con este modelo simplificado estamos reduciendo el número de centros que interaccionan hasta un 6 % del total. Como además el número de interacciones que debemos calcular en la simulación crece con el número de centros de interacción, N , la simplificación en el cálculo es muy superior al 6 % mencionado. Con nuestro modelo el número de interacciones a calcular crece en función del número de centros de interacción como N^2 . Por tanto, al utilizar el modelo *Coarse Grained* el cálculo se reduce hasta un $0.36 \% = (6 \%)^2$ del total de interacciones calculadas con el modelo *full-atom*. De esta manera queda plasmado cómo la utilización de un modelo reducido se hace necesaria para poder hacer ciertos cálculos, que de hacerse a nivel atómico serían muy difíciles de llevar a cabo, al menos con un coste computacional moderado.

Es importante recordar que la utilización de los modelos *Coarse Grained* implica una disminución de la resolución en nuestros resultados. De alguna forma se perderá, como es lógico, parte de la información que sí obtendríamos en caso de hacer simulaciones con modelos no simplificados, Figura 2.4. Sin embargo para nosotros no supone un problema grave, puesto que para el estudio de la termodinámica de un proceso tan

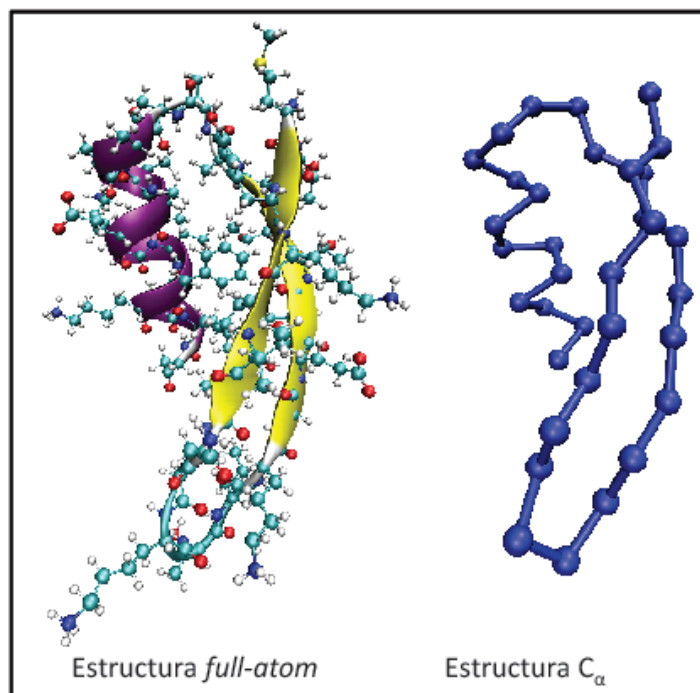


Figura 2.4: **Comparación de la estructura atómica y de la estructura reducida de un fragmento de proteína.** Se muestra la comparación entre una estructura atómica donde se representa cada átomo de la proteína, y la estructura de C_{α} donde solo hay un centro de interacción por residuo. En la imagen de la estructura *full atom* se muestra una representación esquemática de la estructura secundaria, y los distintos tipos de átomos en diferentes colores. Ambas estructuras están representadas en idéntica perspectiva, para que su comparación sea más sencilla.

complejo como es el plegamiento/desnaturalización de las proteínas, esta simplificación no implica una pérdida radical de resolución sino que nos recuerda, más que nunca, que nuestro modelo es una aproximación a la realidad. Debemos ser conscientes en cada momento de lo que se le puede exigir a nuestro modelo.

Si bien en esta Sección hemos justificado el por qué elegimos utilizar un modelo simplificado de tipo *Coarse Grained* en vez de un modelo de más resolución tipo *full-atom*, en la próxima Sección explicaremos las aproximaciones físicas y geométricas que debemos tener en cuenta para que nuestro modelo reproduzca de la forma más exacta posible los tránsitos conformacionales que se dan en la cadena polipeptídica.

2.2.1. Potencial de interacción

Una vez seleccionados los centros sobre los que vamos a aplicar las interacciones (como hemos dicho anteriormente será en los $C\alpha$), el siguiente paso consiste en definir el potencial de interacción que aplicaremos entre estos centros, para que nuestra cadena polipéptica reproduzca, al menos de forma aproximada, las interacciones que soportan los distintos aminoácidos de las proteínas en el proceso de plegamiento. Nuestro grupo de investigación lleva años utilizando los denominados potenciales de tipo $G\bar{o}$ ^{112,113}, desarrollando distintos modelos para simular el plegamiento de proteínas con este tipo de potencial^{114,115}. Los potenciales de tipo $G\bar{o}$, o potenciales basados en la estructura nativa, han sido descritos ampliamente en la bibliografía como un potencial sencillo y de bajo coste computacional, con los que se obtienen buenos resultados, especialmente cuando se utilizan en modelos de tipo *Coarse-Grained*¹¹⁶. Por ejemplo, los trabajos de Prieto y col. muestran el desarrollo y las optimizaciones que se han llevado a cabo en un potencial de este tipo para poder estudiar el plegamiento de proteínas a presión atmosférica^{117,118}. Según el principio de frustración mínima en el que se basan este tipo de potenciales, son las interacciones del estado nativo las que dirigen el plegamiento. Se define un contacto nativo como un par de residuos que se encuentran en contacto en el estado nativo de la proteína, y serán sólo a estos pares a los que tengamos que aplicar el potencial de interacción en cualquier conformación.

En el modelo de Prieto y col.^{102,114}, se definió un potencial de interacción entre contactos nativos según la siguiente ecuación:

$$u_{ij}(r_{ij}) = \begin{cases} (r_{ij} - d_{ij}^{nat})^2 - a^2, & d_{ij}^{nat} - a \leq r_{ij} \leq d_{ij}^{nat} + a \\ 0, & \text{en cualquier otro caso} \end{cases} \quad (2.1)$$

Este potencial, muy sencillo y computacionalmente económico, tiene forma de parábola truncada, como podemos observar en la Figura 2.5. En el eje x se representa

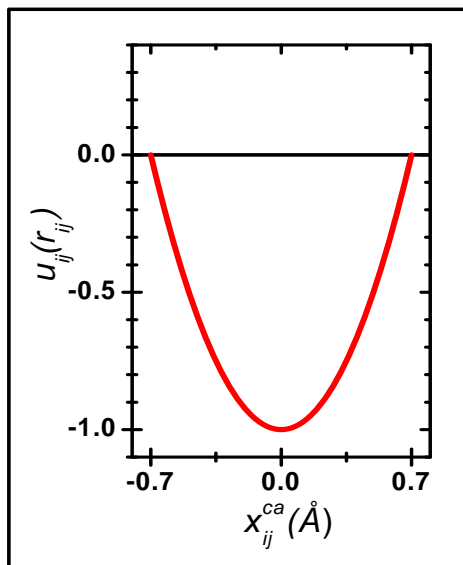


Figura 2.5: **Potencial de interacción tipo Gō**. Representación gráfica del potencial de interacción correspondiente a la ecuación 2.1, que se viene utilizando en nuestro grupo de investigación para estudiar el plegamiento de proteínas a presión atmosférica.

$x_{ij}^{ca} = r_{ij} - d_{ij}^{nat}$, siendo r_{ij} la distancia entre los aminoácidos i y j , y d_{ij}^{nat} la distancia entre esos dos mismos aminoácidos en la estructura nativa de la proteína, medidas ambas entre sus C α . Por tanto, la energía de cada una de las conformaciones de la cadena polipeptídica depende de la distancia de algunos pares de aminoácidos con respecto a la distancia de esos mismos pares en la conformación nativa de la proteína. De esta manera, el mínimo de energía lo encontramos cuando $r_{ij} = d_{ij}^{nat}$, lo que corresponde a la estructura nativa de la proteína.

Además los potenciales de tipo Gō, al ser un tipo de **potenciales de campo medio** o *Potencial of Mean Force* (PMF)¹¹⁹, plantean una ventaja adicional, y es que no tienen en cuenta de forma explícita las moléculas de disolvente, sino que mantienen su efecto de forma implícita, de tal forma que el propio potencial recoge el comportamiento promedio de la interacción entre aminoácidos en un disolvente acuoso¹²⁰. Esto supone una gran ventaja en la simulación, puesto que no produce gasto computacional adicional debido a la consideración detallada del disolvente. Al mismo tiempo, tampoco existe una pérdida excesiva de información en los resultados obtenidos.

Bajo esa premisa, en esta Tesis planteamos la posibilidad de desarrollar un poten-

cial de interacción con propiedades similares a las que tienen los potenciales de tipo $G\bar{o}$, que nos sirva para describir el comportamiento del sistema a diferentes presiones. Los potenciales de tipo $G\bar{o}$ son potenciales basados en la estructura del estado nativo, lo que significa que es necesario conocer el estado nativo de la proteína que se va a estudiar para poder aplicar este tipo de potencial. Con el modelo **en ningún caso pretendemos predecir el estado nativo de una proteína**, sino estudiar el proceso de plegamiento entre un estado nativo conocido y estados intermedios o desnaturalizados que sí son desconocidos.

Como se describió en la Introducción, el efecto que ejerce la presión en el proceso de plegamiento afecta principalmente a la interacción proteína-disolvente y no tanto a la propia cadena covalente. Esto podría suponer un problema para nuestra simulación, puesto que como ya hemos comentado los potenciales de tipo $G\bar{o}$ no tienen en cuenta la presencia del disolvente de forma explícita. Pero si conseguimos un PMF que reproduzca el efecto del disolvente, podremos obtener un potencial de interacción que mantenga de forma implícita el efecto del medio a pesar de no usar moléculas de disolvente en nuestra simulación.

En este punto cabe recordar que las interacciones intramoleculares más importantes que rigen el plegamiento de proteínas son las interacciones que se producen entre los distintos grupos hidrófobos de los aminoácidos¹²¹. Por esta razón, la mayoría de los potenciales de campo medio que se utilizan para simular el plegamiento de proteínas se basan en la reproducción de las fuerzas que se generan entre centros hidrófobos en un medio acuoso. Existen diferentes potenciales de campo medio que muestran el perfil de energía en función de la distancia entre residuos hidrófobos en agua¹²²⁻¹²⁴. En estos potenciales observamos siempre dos pozos de interacción atractiva separados por una barrera de repulsión. En la Figura 2.6 (a) podemos observar una imagen de este tipo de potencial. Esta imagen está tomada del trabajo de Chan y col¹²⁵. El pozo de interacción a mayor distancia coincide con la interacción de dos moléculas de metano cuando se

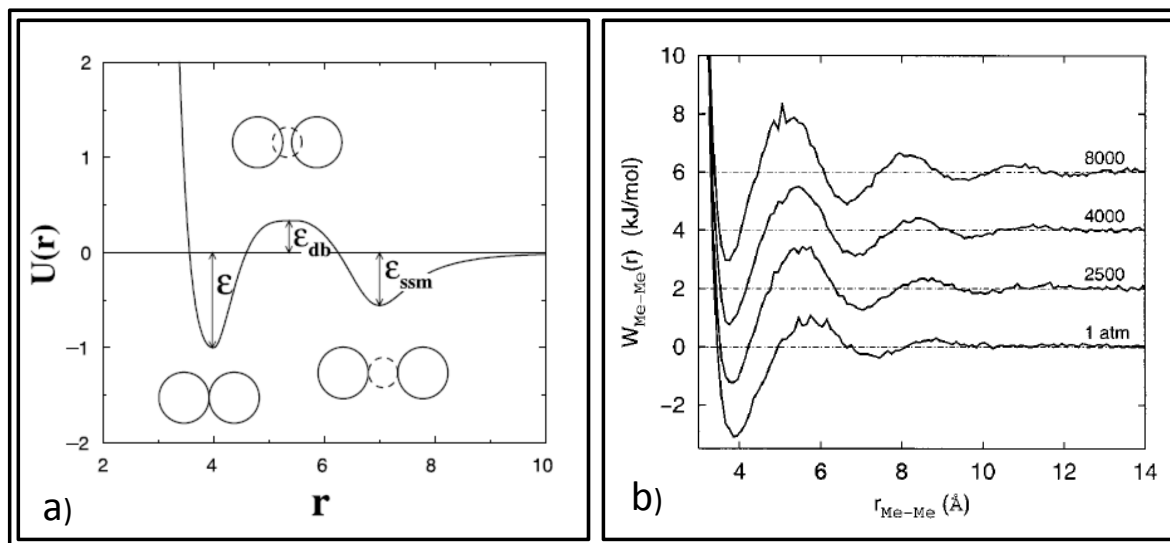


Figura 2.6: **Potenciales de campo medio recogidos en la bibliografía:** a) Esta imagen, tomada de¹²⁵, muestra un potencial de campo medio para dos moléculas de metano en agua a presión atmosférica. En el dibujo esquemático las moléculas de metano se representan por las circunferencias de línea continua, mientras que la molécula de agua se representa por una circunferencia con línea discontinua. Se pueden observar los dos pozos de interacción separados por una barrera de desolvatación. b) Esta imagen, tomada de¹²⁶, muestra los diferentes potenciales de campo medio, obtenidos por dinámica molecular, para dos moléculas de metano a presiones crecientes. Se muestran los perfiles para 1, 2500, 4000 y 8000 atm. Se puede observar cómo a medida que aumenta la presión, la profundidad de los pozos y la altura de la barrera cambian, observándose una relación entre sus profundidades y alturas y la presión del medio.

encuentran separadas por una molécula de agua; el pozo de interacción a menor distancia coincide con la interacción de dos moléculas de metano directamente entre ellas, sin molécula de disolvente entre medias; la barrera de repulsión que se forma corresponde con la expulsión de la molécula de disolvente que separa las dos moléculas de metano cuando estas se van acercando.

Existen también estudios con potenciales obtenidos a través de simulación molecular a distintas presiones^{126–128}. Estos estudios demuestran que la interacción entre residuos hidrófobos dentro de un disolvente acuoso va a ir variando en función de la presión a la que esté sometido el sistema. En la Figura 2.6 (b), que hemos tomado del trabajo de Garde y col.¹²⁶, se puede observar cómo los potenciales de campo medio para dos moléculas de metano en agua son diferentes a las distintas presiones que se presentan. Existe una evolución de la interacción neta entre las dos moléculas hidrófobas en función de la presión hidrostática a la que están sometidas.

En esta misma Figura podemos observar que aunque los potenciales de interacción evolucionan en función de la presión del sistema, existen características comunes a estos potenciales de campo medio en el intervalo de presiones que se muestra. Observamos cómo se mantienen los distintos pozos de interacción separados por la barrera repulsiva: aunque la anchura de estos pozos y barreras apenas cambia con la presión, sí que se observa una evolución en la altura de la barrera y la profundidad de los pozos en función de la presión. De tal forma que a mayor presión, más altas son las barreras y especialmente más profundos son los pozos de interacción, fundamentalmente el situado a mayor distancia.

A partir de las dos ideas planteadas, la existencia de potenciales de campo medio con los que se puede reproducir el efecto de interacciones hidrófobas en agua y el estudio de cómo varían estos en función de la presión del medio, nosotros aprovechando la experiencia previa del grupo con los potenciales de tipo $G\bar{o}$ hemos planteado en esta Tesis nuestro propio potencial de interacción, que se muestra en la Figura 2.7. Este potencial mantiene dos pozos de interacción separados por una barrera repulsiva.

El primer pozo de interacción corresponde a la interacción directa entre residuos que forman un contacto nativo. Lo hemos denominado **pozo $G\bar{o}$** , por su semejanza con los potenciales de interacción de tipo $G\bar{o}$ que nuestro grupo venía utilizando (Figura 2.5). El segundo pozo lo denominamos **pozo ssm** , del inglés, *solvent separated minimum*. Entre ambos encontramos una barrera de desolvatación, la **barrera db** , del inglés *desolvation barrier*.

La definición matemática del potencial de interacción entre pares de residuos ij , que forman un contacto nativo en la estructura nativa, es la siguiente:

$$u_{ij}(r) = \begin{cases} \frac{\epsilon_{go}}{a^2} \left[(r - d_{ij}^{nat})^2 - a^2 \right], & d_{ij}^{nat} - a < r \leq d_{ij}^{nat} + a \\ -\frac{\epsilon_{db}}{b^2} \left[(r - d_{ij}^{nat} - a - b)^2 - b^2 \right], & d_{ij}^{nat} + a < r \leq d_{ij}^{nat} + a + 2b \\ \frac{\epsilon_{ssm}}{a^2} \left[(r - d_{ij}^{nat} - 2a - 2b)^2 - a^2 \right], & d_{ij}^{nat} + a + 2b < r \leq d_{ij}^{nat} + 3a + 2b \\ 0, & \text{en cualquier otro caso} \end{cases} \quad (2.2)$$

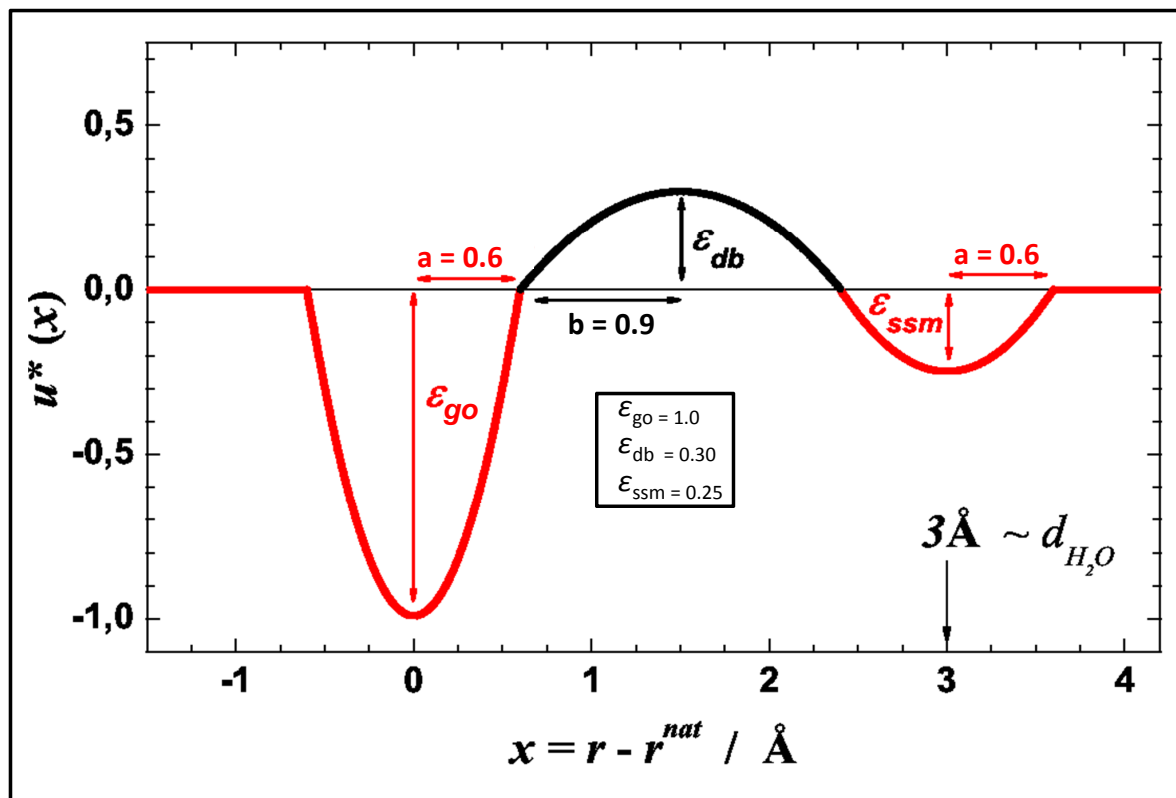


Figura 2.7: **Potencial de interacción desarrollado en esta Tesis.** Mostramos el potencial de interacción que vamos a aplicar para los residuos que participan en contactos nativos. El potencial está compuesto por dos parábolas de interacción atractiva, separadas por una parábola que hace de barrera repulsiva. El valor del parámetro a es fijo, 0.6 \AA , y por geometría el parámetro b vale 0.9 \AA , ya que la distancia entre los dos pozos es de 3 \AA . El potencial se puede modular en función del valor que tomen los parámetros ϵ_{go} , ϵ_{db} y ϵ_{ssm} , para así poder definir distintas presiones.

donde r corresponde a la distancia entre los residuos i y j , y d_{ij}^{nat} corresponde a esta misma distancia para la estructura nativa de la proteína. Tanto los pozos de interacción como la barrera de desolvatación están descritos matemáticamente como una parábola, con signo negativo y anchura a en caso de los pozos $G\bar{o}$ y ssm , y signo positivo y anchura b en el caso de la barrera db . El primer pozo de interacción $G\bar{o}$, truncado en $\pm a$, está centrado en $x = 0$. Recordemos que definimos x como la diferencia de distancias que existe para cada par de aminoácidos entre r_{ij} y d_{ij}^{nat} . El centro del pozo ssm se sitúa a 3 \AA del pozo $G\bar{o}$. Estos 3 \AA corresponden al diámetro considerado habitualmente para una molécula de agua¹²⁵. Este segundo pozo atractivo está truncado en $\pm a$ respecto a su posición de equilibrio. Por último, la barrera de desolvatación (db) corresponde a

una función parabólica situada en el punto medio entre los dos pozos, y truncada en un valor de $\pm b$. El valor del parámetro a lo hemos mantenido en 0.6 \AA en función del análisis previo sobre la influencia de este parámetro en el proceso de plegamiento a presión atmosférica¹¹⁴ y, por consiguiente, por geometría el valor del parámetro b es igual a 0.9 \AA .

La expresión matemática del potencial y la Figura 2.7 no muestran la contribución repulsiva debida al volumen excluido. En nuestro modelo los $C\alpha$ se consideran como esferas rígidas, por lo que se prohíben todas las conformaciones donde la distancia entre pares de aminoácidos sea inferior a 4.2 \AA , valor que previamente ha sido ajustado en nuestro grupo de investigación¹¹⁴. Para el modelo desarrollado en este trabajo hemos tenido en cuenta una consideración añadida: para que la cadena polipéptica pueda ocupar a lo largo de las simulaciones las cavidades internas que pueda tener la proteína en su estado nativo, hemos decidido que en los casos donde dicha distancia de corte quede situada en el interior de la parábola del pozo $G\bar{o}$, la desplazaremos hasta $d_{ij}^{mat} - a$, de tal forma que el pozo $G\bar{o}$ queda plenamente accesible para el par de residuos interaccionantes. De esta manera conseguimos que con nuestro modelo se puedan reproducir configuraciones que en ciertas zonas pudieran estar más empaquetadas localmente que la proteína en el estado nativo.

Además, es importante mencionar que en este trabajo hemos introducido una barrera repulsiva que está desacoplada con respecto a los pozos de interacción. No es posible llevar a cabo este proceso con potenciales de tipo Leonnard-Jones. Por ejemplo, existen trabajos recientes que muestran cómo este desacoplamiento produce unos resultados sensiblemente mejores para el plegamiento de proteínas con potenciales basados en la estructura.^{129,130}

Los parámetros más importantes del modelo son la profundidad de los pozos de interacción y la altura de la barrera db . Corresponden a los parámetros ϵ_{go} , ϵ_{db} y ϵ_{ssm} , indicados en la Figura 2.7. En nuestro modelo, como en otros modelos desarrollados por

el grupo, hemos fijado ϵ_{go} como $1/a^2$, de modo que el mínimo de energía de cada interacción de este tipo corresponde a un valor de -1 en unidades reducidas, ecuación (2.2), de tal forma que cuando $\epsilon_{db} = \epsilon_{ssm} = 0$, el único mínimo de energía corresponde al estado nativo. En este caso se recupera el potencial de la Figura 2.5. Los parámetros ϵ_{db} y ϵ_{ssm} los normalizamos con respecto a ϵ_{go} , para así tener todos los parámetros relacionados entre sí. La presión que apliquemos al sistema dependerá por tanto de la relación que exista entre estos tres parámetros:

$$\text{Presión de sistema} \sim (\epsilon_{db}/\epsilon_{go} ; \epsilon_{ssm}/\epsilon_{go}) \quad (2.3)$$

siguiendo la idea de los resultados recogidos en la Figura 2.6 (b).

2.2.2. Mapa de contactos

La manera de definir los contactos presentes en el estado nativo admite diferentes posibilidades. Por esa razón decidimos mantener el mismo criterio que viene usando nuestro grupo de investigación para otros modelos^{102,114}. Hemos seleccionado una distancia de corte (d_{max}) de 4.5 Å; todas las parejas de residuos que, en el estado nativo, tengan átomos pesados (átomos que no sean H) a una distancia inferior a d_{max} formarán un contacto, y sólo sobre estos contactos aplicaremos nuestro potencial de interacción. Cualquier interacción entre C α por otra parte, está sujeta a las restricciones de volumen excluido.

Para poder visualizar fácilmente qué C α forman o no contactos nativos utilizamos los denominados **mapas de contactos**. En estos mapas se representa la secuencia de aminoácidos en el eje de abscisas y en el eje de ordenadas. Cada punto presente en el mapa muestra que ese par de aminoácidos i,j se encuentra formando un contacto nativo, y por tanto será a esta pareja de aminoácidos a la que le apliquemos el potencial de interacción de la ecuación (2.2). La Figura 2.8 muestra el mapa de contactos que hemos obtenido para un fragmento de una proteína. Se puede observar por ejemplo cómo los residuos 7 y 33, que por secuencia no se encuentran cercanos y sin embargo espacialmente

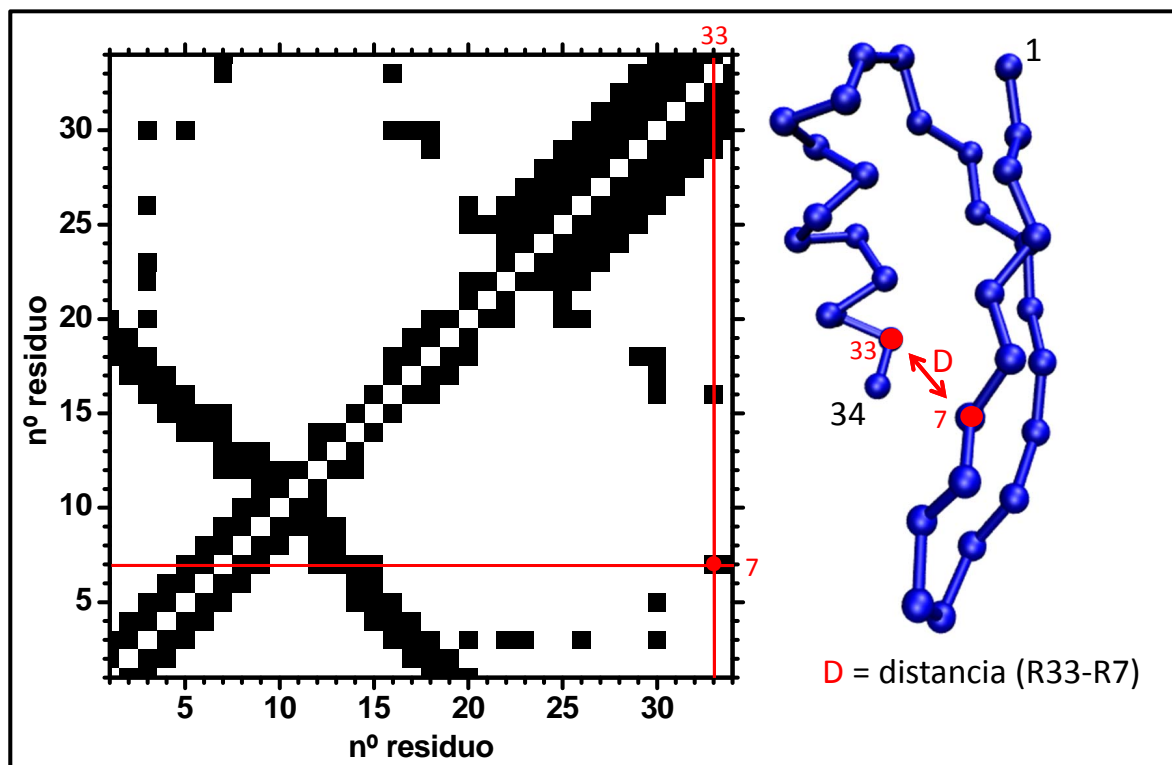


Figura 2.8: **Mapa de contactos nativos de un fragmento de proteína.** Mapa de contactos generado para el mismo fragmento proteico de la Figura 2.4. Cada punto en el mapa señala que dos residuos están formando un contacto nativo. En el ejemplo que mostramos, en rojo podemos ver cómo los residuos 7 (R7) y 33 (R33) forman un contacto nativo.

sí lo están, forman un contacto en su estructura nativa. De esta manera, el mapa de contactos nos permite traducir, de una forma muy sencilla, la estructura tridimensional de la proteína a una representación bidimensional mucho más fácil de analizar. Además, es la base para la definición del potencial de interacción del modelo.

Para definir nuestro modelo adecuadamente en un estudio de la influencia de la presión, hemos dividido los contactos del estado nativo en tres tipos de contactos, dependiendo de su naturaleza y del potencial de interacción que vayamos a aplicarles: contactos locales, contactos de enlace de hidrógeno y contactos de largo alcance.

Contactos locales:, se trata de contactos debidos a la conectividad de la cadena. Para los pares de aminoácidos (i,j) donde $|i - j| < 4$, definiremos siempre la existencia de un contacto nativo local, independientemente del criterio de distancias mencionado anteriormente. Los residuos cercanos en secuencia forman interacciones de ángulos de

enlace virtual entre los aminoácidos i e $i + 2$, e interacciones de ángulos de torsión virtual entre los aminoácidos i e $i + 3$. Precisamente estas última interacciones son las responsables de la quiralidad de la proteína en nuestro modelo¹⁰². Los contactos locales son importantes para la formación de la estructura secundaria. Para este tipo de contactos sólo aplicaremos la primera parte del potencial de interacción, la correspondiente al pozo $G\bar{o}$. Por tanto para estos contactos tomamos $\epsilon_{db} = \epsilon_{ssm} = 0$. Se trata de los contactos en la cadena principal de la proteína. Estos aminoácidos se encuentran tan cercanos que físicamente es muy difícil que una molécula de disolvente se sitúe entre ellos, por esa razón no aplicamos la parte del potencial relacionada con de la presión y nos limitamos a aplicar un potencial de tipo $G\bar{o}$, como el que utilizamos a presión atmosférica.

Contactos de enlace de hidrógeno:, se trata de contactos entre residuos lejanos en secuencia, $|i - j| \geq 4$, pero que forman enlaces de hidrógeno en la estructura nativa. Estos enlaces de hidrógeno son, junto a las interacciones hidrófobas, las principales fuerzas que mantienen las estructuras secundarias de las proteínas, como se muestra en trabajos recientes de nuestro grupo de investigación^{131,132}. Para este tipo de contacto, al igual que sucediera con los contactos locales, sólo aplicamos el pozo $G\bar{o}$ de nuestro potencial de interacción. De nuevo, dada la naturaleza parcialmente covalente de un enlace de hidrógeno, no es razonable considerar que una molécula de agua se introduzca entre los átomos dador y aceptor del mismo manteniendo la interacción. La selección de los contactos de hidrógeno en el estado nativo se realizó con el programa de simulación Pymol¹³³, que gracias a la posición de todos los átomos es capaz de calcular qué aminoácidos son los que forman enlaces de hidrógeno en la estructura secundaria.

Contactos de largo alcance, se trata de contactos entre residuos que no son cercanos en secuencia, $|i - j| \geq 4$, y que además no forman enlace de hidrógeno. Será a estos contactos a los que apliquemos el potencial completo de la ecuación (2.2) con valores de $\epsilon_{db} \neq 0$ y $\epsilon_{ssm} \neq 0$. Por tanto, será a este tipo de contactos a los que apliquemos

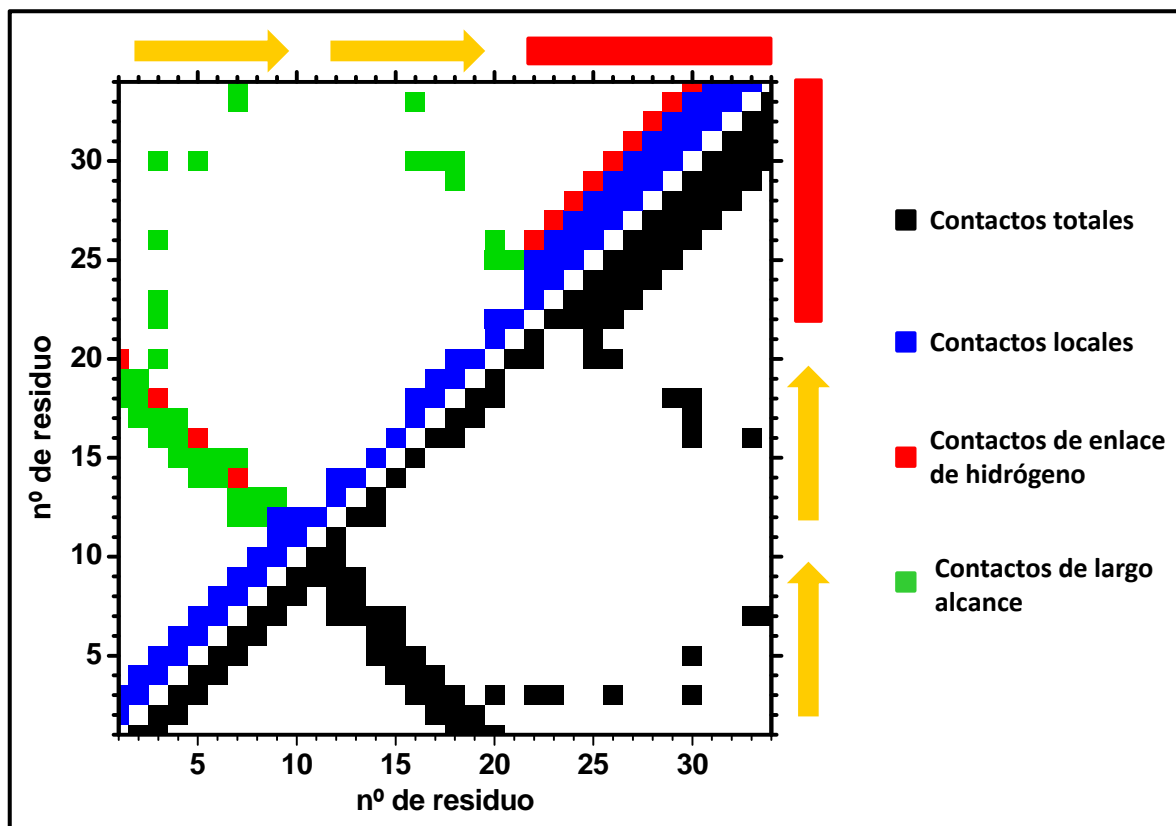


Figura 2.9: **Mapa de tipos de contactos nativos para un fragmento de proteína.** Mostramos el mapa de contactos completo, donde se clasifican los distintos tipos de contacto existentes. El mapa está dividido en dos partes: en el triángulo inferior-derecho encontramos todos los contactos, este triángulo es exactamente igual a la Figura 2.8. El triángulo superior-izquierdo muestra los mismos contactos divididos en los tres posibles tipos de contacto: en azul los contactos locales, en rojo los contactos de enlace de hidrógeno y en verde los contactos de largo alcance. Las flechas y rectángulos en el exterior del mapa, en amarillo y rojo, muestran los elementos de estructura secundaria, (hebra β y hélice α respectivamente).

nuestro potencial de interacción de alta presión, ya que precisamente entre estos residuos más alejados en la secuencia es donde físicamente se puedan introducir las moléculas de agua que llegan desde el medio que rodea a la proteína debido al aumento de la presión.

En la Figura 2.9 podemos observar el mapa con los diferentes tipos de contactos para el mismo fragmento proteico que la Figura 2.8. El mapa que presentamos está dividido en dos partes: en el triángulo inferior derecho mostramos en negro todos los contactos existentes en la estructura nativa, y nos sirve de referencia para así saber el número total de contactos que tiene la proteína que estudiamos. En el triángulo superior-izquierdo mostramos los mismos contactos pero esta vez diferenciados por el tipo de contacto que forman. En azul se muestran los contactos locales, en rojo los

contactos de enlace de hidrógeno y por último en verde se muestran los contactos de largo alcance, a los que aplicaremos el potencial completo. La información existente en esta representación detallada del mapa de contactos es la que usaremos en este trabajo para realizar las simulaciones. Además, para poder visualizar mejor la estructura de la proteína a partir de su mapa de contactos, usamos los esquemas que hay junto a los ejes: las flechas muestran estructura β en la proteína mientras que los rectángulos representan la estructura α . De esta manera esquemática se hacen más visibles los motivos de estructura secundaria de la proteína.

2.3. Método de simulación

Una vez descrito el modelo geométrico, debemos describir también el método de simulación que hemos utilizado para poder evaluar las propiedades que queremos estudiar en nuestro sistema de interés. Puesto que nuestro objetivo es el estudio de las propiedades termodinámicas y estructurales del plegamiento de proteínas, hemos seleccionado un método que sea capaz de reproducir adecuadamente los estados de equilibrio donde pueden coexistir configuraciones muy diferentes a la misma temperatura. Para ello hemos utilizado el método de Monte Carlo y lo hemos combinado con la técnica de *parallel tempering*.^{134,135} Mediante el uso de esta técnica logramos estudiar el sistema en un intervalo amplio de temperaturas, al mismo tiempo que minimizamos las posibilidades de que el sistema quede atrapado en un mínimo local de energía.

Como adelantamos en la Sección 2.1, el método de Monte Carlo nos proporciona un muestreo aleatorio de la superficie de energía del sistema que deseamos estudiar. Utilizamos además un muestreo de importancia, donde cada una de las configuraciones generadas recibirá una probabilidad de existencia que depende de su factor de Boltzmann, y por tanto de la temperatura y energía del sistema. En nuestro caso, el muestreo de importancia lo incluimos a través del **criterio de Metropolis**.¹³⁶ Los pasos a seguir para poder realizar nuestra simulación son los siguientes:

- Seleccionamos la proteína que deseamos estudiar, y generamos su mapa de contactos para esta proteína a partir de la información existente de su estado nativo. Esta información la hallamos en una base de datos de estructuras experimentales, el *Protein Data Bank* (**PDB**)¹³⁷.
- Seleccionamos una configuración inicial C_0 . Puede ser la configuración nativa de la proteína, o una configuración aleatoria que esté permitida geométricamente. Para esta configuración calculamos su energía E_0 a partir del potencial de interacción que presentamos en la Sección anterior.
- A partir de C_0 generamos una nueva configuración C_1 , de la cual también calcularemos su valor de energía E_1 . Al final de esta misma Sección explicaremos las reglas que utilizamos para generar nuevas configuraciones a partir de las anteriores.
- En este momento aplicamos el criterio de Metropolis, para decidir si la nueva configuración que se ha generado debe de ser considerada o no en el muestreo. Para ello comparamos las dos energías calculadas E_0 y E_1 , encontrando dos posibilidades:
 1. Si $E_1 < E_0$, la nueva configuración C_1 se acepta automáticamente, y sustituye a la anterior configuración C_0 . Al ser la nueva configuración de menor energía, resulta ser más estable que la anterior y por tanto la probabilidad estadística de esta configuración se hace mayor que la anterior a la misma temperatura.
 2. Si $E_1 > E_0$, la probabilidad de que se produzca un cambio de configuración dependerá de la relación de los factores de Boltzmann de ambas configuraciones:

$$p_{01} = \frac{e^{-\beta E_1}}{e^{\beta E_0}} = \exp[-\beta(E_1 - E_0)] = \exp(-\beta \Delta E) \quad (2.4)$$

donde $\beta = 1/k_b T$, siendo k_b la constante de Boltzmann y T la temperatura. Para poder aceptar la nueva configuración de mayor energía con una probabilidad p_{01} se

genera un número aleatorio λ dentro del intervalo $(0,1)$. En caso de que el valor de λ fuese menor a $\exp(-\beta\Delta E)$, el movimiento se acepta y la configuración C_1 sustituye a la configuración C_0 . En caso de que el valor de λ fuese mayor a $\exp(-\beta\Delta E)$, la configuración C_1 se rechaza y al sistema no le dejamos evolucionar manteniéndolo en E_0 . En este caso, la configuración C_0 se cuenta de nuevo en el muestreo y seguirá siendo el punto de partida para generar la próxima configuración.

- Aquí habrá terminado un paso de Monte Carlo y volvemos al primer punto, donde iniciamos de nuevo el bucle, a partir de la configuración que el criterio de Metropolis haya determinado.

En la Figura 2.10 mostramos un diagrama de flujo donde de manera esquemática se pueden ver resumidas todas las etapas que necesitamos para realizar un paso de Monte Carlo.

Aunque tradicionalmente el método de Monte Carlo ha sido aplicado a temperatura constante, o en una serie de temperaturas independientes entre sí, en este trabajo hemos utilizado la técnica del *parallel tempering* o intercambio de réplicas como alternativa para poder estudiar el sistema simultáneamente a varias temperaturas y mejorar el muestreo. Esta técnica consiste en crear N réplicas del sistema a estudiar. Cada una de las réplicas o copias se encuentra a una temperatura diferente, y evoluciona según el esquema de Monte Carlo que acabamos de explicar. A cada temperatura la probabilidad de aceptación o rechazo de cada configuración depende del criterio de Metropolis y por tanto de la energía del sistema y la temperatura de la réplica. Tras una cantidad estipulada de pasos de Monte Carlo se produce un intercambio de configuraciones entre temperaturas consecutivas, de tal forma que de nuevo mediante un criterio de Metropolis, y por tanto dependiendo de la diferencia de energías entre dos configuraciones a diferentes temperaturas, se acepta o no el intercambio de réplicas, es decir, que la configuración C_0 de energía E_0 a T_0 se intercambie por la configuración C_1 de energía E_1 a T_1 .

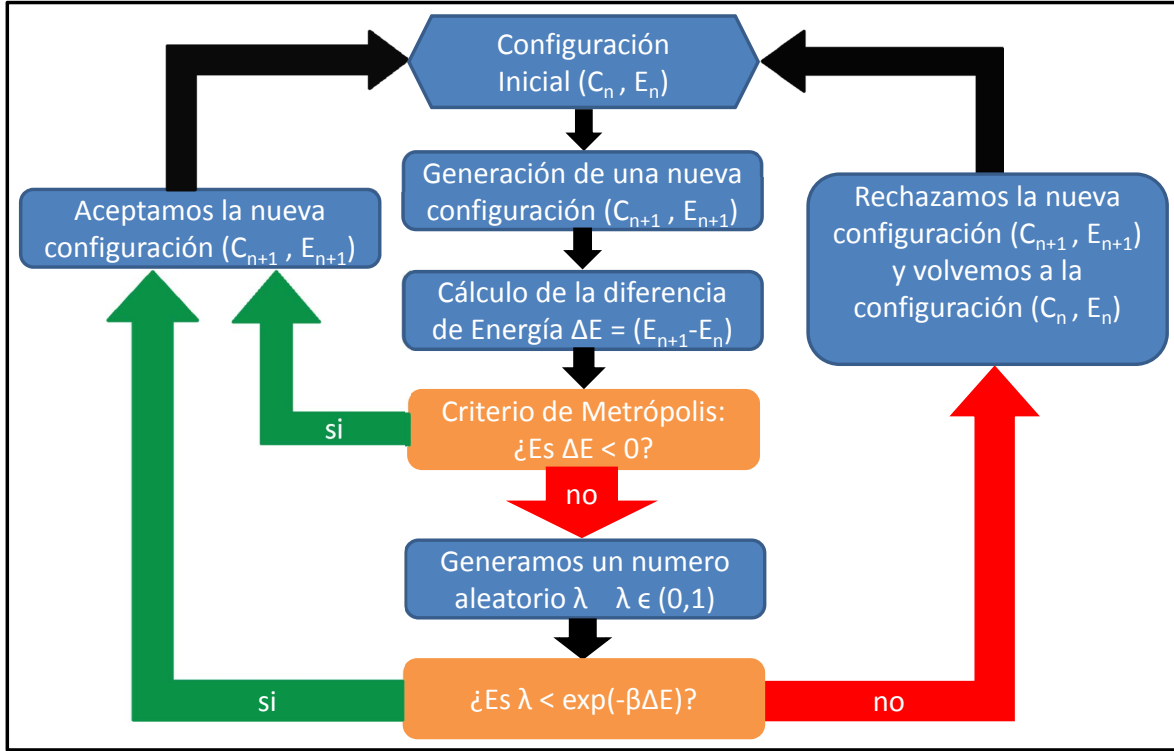


Figura 2.10: **Diagrama de flujo para un paso de Monte Carlo.** Mostramos mediante un diagrama de flujo las diferentes etapas a seguir para poder realizar un paso de Monte Carlo en nuestro modelo.

A la temperatura T_0 la probabilidad de intercambio es, según la ecuación (2.4):

$$p_{01} = \exp[-\beta_0(E_1 - E_0)] \quad (2.5)$$

Del mismo modo, a la temperatura T_1 la probabilidad de intercambio es:

$$p_{10} = \exp[-\beta_1(E_0 - E_1)] \quad (2.6)$$

Y por tanto la probabilidad de que se intercambien entre sí ambas configuraciones será igual a:

$$p_{intercambio} = p_{01} \times p_{10} = \exp[-\beta_0(E_1 - E_0)] \exp[-\beta_1(E_1 - E_0)] \quad (2.7)$$

que podemos expresar como:

$$p_{intercambio} = \exp[(\beta_1 - \beta_0)(E_1 - E_0)] = \exp(\Delta\beta\Delta E) \quad (2.8)$$

De nuevo, generaremos un número aleatorio λ entre 0 y 1; en caso de que λ sea inferior a $\exp(\Delta\beta\Delta E)$, el intercambio de réplicas entre C_0 y C_1 será aceptado; en caso de que λ sea superior a $\exp(\Delta\beta\Delta E)$ rechazaremos el intercambio de réplicas. Por tanto, para que se pueda aceptar el intercambio es necesario que las configuraciones que deseamos intercambiar tengan una energía relativamente parecida, y por eso los intercambios de réplicas se intentan entre réplicas a temperaturas consecutivas, donde las configuraciones son más similares entre sí. Esto implica hacer una buena selección de las temperaturas a las que realizamos la simulación, ya que queremos seleccionar el mínimo número de temperaturas, para poder así tener el menor coste computacional posible, pero al mismo tiempo tiene que existir un solapamiento efectivo entre las distribuciones de energía de las temperaturas vecinas para que los intercambios sean aceptados en una proporción razonable. Por esta razón, el número de temperaturas a las que estudiamos cada proteína dependerá del tipo de tránsito que sufra la proteína en cuestión. Tránsitos más abruptos conllevarán un mayor número de temperaturas, mientras que tránsitos más continuos necesitarán un menor número de temperaturas para su estudio.

Como ya comentamos, esta técnica no sólo nos permite estudiar el sistema paralelamente a distintas temperaturas, sino que además permite un muestreo mucho más eficiente puesto que el intercambio de réplicas hace que todas las réplicas pasen por cada una de las temperaturas numerosas veces a lo largo de la simulación, haciendo mucho más complicado que el sistema quede atrapado en mínimos locales, ya que aunque una configuración pueda quedar atrapada en un mínimo local a baja temperatura, al intercambiarse a una temperatura superior el sistema tendrá la posibilidad de salir del mínimo de energía. Precisamente el quedar atrapados en mínimos locales y no hacer un muestreo eficiente del embudo de plegamiento es uno de los mayores inconvenientes del

método de Monte Carlo clásico a una única temperatura¹³⁸.

Un aspecto importante en nuestras simulaciones son los movimientos que aplicamos a la cadena polipeptídica para poder realizar un buen muestreo de la superficie de energía. Ya dijimos al principio de esta Sección que definimos una serie de movimientos aleatorios, que generan configuraciones que serán aceptadas o rechazadas en función de los criterios ya explicados. Para nuestro modelo los movimientos que efectuamos sobre la cadena de residuos son de tres tipos¹⁰²:

1. **Movimiento de final de cadena:** este tipo de movimiento se aplica al primero y al último de los $C\alpha$ de la cadena polipeptídica. Consiste en generar un nuevo vector de enlace virtual entre el $C\alpha$ terminal y el $C\alpha$ vecino, definiendo así una nueva posición en el principio o final de la cadena. En la Figura 2.11 (a) podemos observar este movimiento para el primer $C\alpha$ de una cadena polipeptídica.
2. **Movimiento de tipo ‘*spike*’:** este movimiento, que se aplica a las unidades interiores de la cadena, conlleva un movimiento de rotación del $C\alpha$ del residuo i alrededor del eje que une a los $C\alpha$ de los residuos $i-1$ e $i+1$. Con el movimiento de *spike* conseguimos, entre otras cosas, que residuos individuales ocupen espacios accesibles en cavidades internas la proteína. En la Figura 2.11 (b) podemos observar este movimiento en el residuo número 4; no sólo se genera una nueva posición en este residuo sino que además conlleva la generación de dos nuevos vectores, v_3 y v_4 .
3. **Movimiento interno con desplazamiento:** este movimiento también se aplica a las unidades internas de la cadena. Se trata de un movimiento de final de cadena pero para un residuo interno, seguido de un desplazamiento del resto de la cadena. Dado el $C\alpha$ del residuo i , se genera un nuevo vector de enlace virtual entre el $C\alpha$ del residuo $i-1$ y el $C\alpha$ del residuo i ; a continuación el resto de la cadena sufre un desplazamiento análogo al del aminoácido i , de tal forma que ningún otro vector

[h]

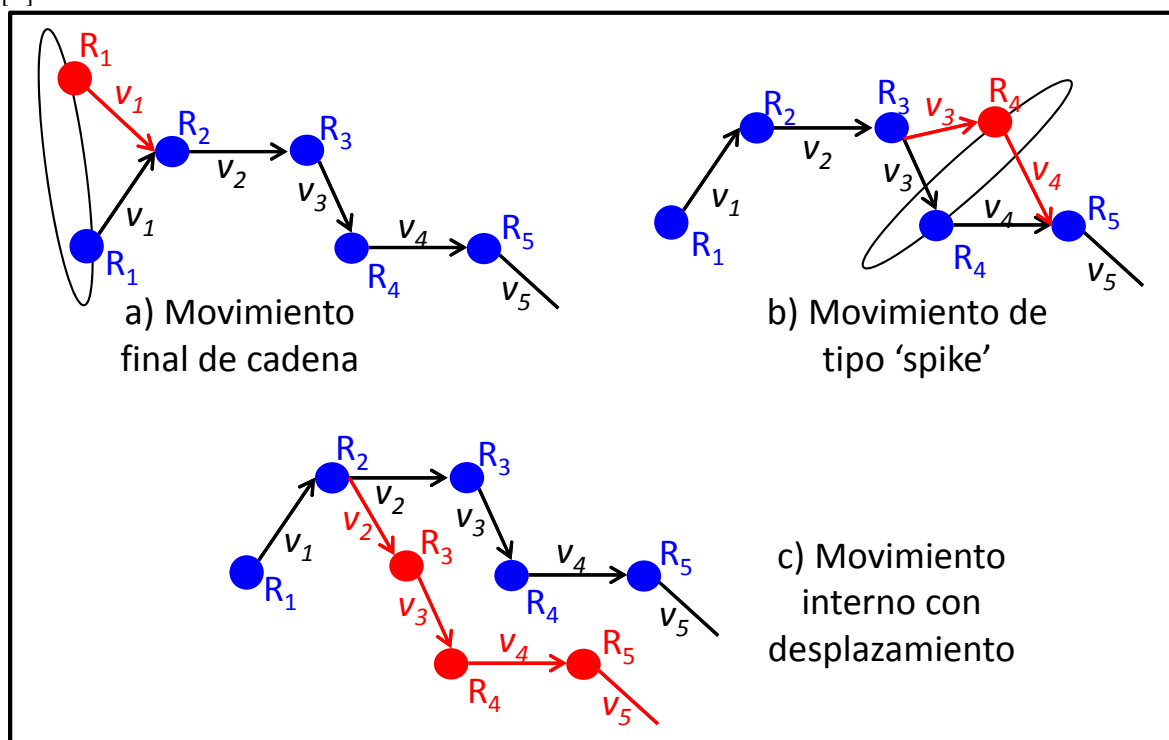


Figura 2.11: **Movimientos en la cadena polipeptídica.** Mostramos los diferentes movimientos que se pueden dar con nuestro modelo. Los círculos representan residuos (R) y las flechas representan vectores de enlace (v). El color azul y negro muestra la posición de los residuos y vectores antes del movimiento, en color rojo se muestran los residuos y vectores tras el movimiento.

de enlace se modifica. Con este tipo de movimiento logramos desplazamientos de mayor calado en las nuevas configuraciones, ya que pueden ser muchos los aminoácidos que se desplazan con este tipo de movimiento aunque con una amplitud limitada. En la Figura 2.11 (c) podemos observar este tipo de movimiento en el residuo número 3. En este caso se genera también una nueva posición para todos los residuos posteriores al número 3, y del mismo modo se genera un nuevo vector número 2, y el desplazamiento de todos los vectores posteriores, aunque manteniendo su orientación.

Después de cada movimiento, y puesto que por lo menos un aminoácido habrá cambiado de posición, se debe comprobar si existen o no solapamientos entre los distintos $C\alpha$. En caso de que no existan se procede al cálculo de la energía de esa configuración y al proceso de aceptación o rechazo del movimiento según el criterio de Metropolis explicado

previamente. Cada nueva conformación generada, ya sea aceptadas o rechazadas, la denominamos **paso de Monte Carlo**. Al primer y último aminoácido les aplicaremos el movimiento 1, mientras que para cada residuo interno seleccionaremos al azar, con un 50 % de probabilidad, si aplicamos el movimiento 2 o el movimiento 3.

Dada una proteína de N aminoácidos, cuando hayamos intentado N movimientos, uno por aminoácido, consideraremos que hemos realizado un **ciclo de Monte Carlo**. Cada ciclo de Monte Carlo está por tanto compuesto por N pasos de Monte Carlo. Nuestras simulaciones conllevan millones de ciclos de Monte Carlo. Además, antes de registrar ningún tipo de resultados, realizamos un número importante de ciclos de termalización, de tal forma que conseguimos que cada una de las réplicas se ajuste a la situación de equilibrio propia de su temperatura. Pasados los ciclos de termalización empezamos a registrar los resultados. Debido al elevado número de ciclos de Monte Carlo que realizamos, sólo registramos las propiedades del sistema en uno de cada 1000 ciclos de Monte Carlo, mientras que el registro de la estructura lo realizamos en uno de cada 10000 ciclos de Monte Carlo. Además, para cada uno de los sistemas estudiados hemos realizado entre 5 y 9 simulaciones independientes a partir de distintas configuraciones iniciales. Los resultados que se muestran en esta memoria son siempre un promedio sobre el total de simulaciones. De esta manera aseguramos la reproducibilidad de nuestra simulación, al mismo tiempo que aumentamos la fiabilidad estadística de los resultados.

2.4. Análisis de Datos

De cada una de nuestras simulaciones obtenemos diferentes archivos con distinto tipo de información, tanto acerca de detalles de la simulación que nos permiten saber si ha podido existir algún tipo de error a lo largo de la misma, como de los resultados registrados sobre las configuraciones que se han obtenido a lo largo del muestreo. Como dijimos en la Sección anterior, existen dos archivos principales, en uno de ellos quedan registradas las propiedades instantáneas del sistema y en otro las estructuras que se han

ido generando.

Información estructural: se trata de un archivo donde en coordenadas cartesianas, **formato .xyz**, se registra cada una de las posiciones de los $C\alpha$ en las conformaciones generadas. De esta manera, gracias a algún programa de visualización molecular (VMD¹³⁹, Makemol¹⁴⁰, Pymol¹³³, etc.) podemos representar de forma sencilla las configuraciones registradas. Así podremos estudiar de manera visual las conformaciones obtenidas, y podemos obtener de su análisis información geométrica del sistema, como por ejemplo las distancias entre los distintos residuos u otras propiedades específicas.

Propiedades del sistema: en este archivo registramos diferentes propiedades globales de cada una de las configuraciones almacenadas. Estas propiedades son: datos de energía de cada configuración, desglosados en función de las diferentes contribuciones, **energía total del sistema**, **energía** debida a los contactos de **corto alcance** y **energía** debida a los contactos de **largo alcance**. Además, en este archivo encontramos datos estructurales de distinto tipo, como por ejemplo los **contactos nativos** presentes en cada una de las conformaciones. Para calcular los contactos nativos que tiene formada cada configuración nos ayudamos del mapa de contactos de la estructura nativa. Consideraremos que existe un **contacto nativo directo** entre cualquier par de residuos i,j que forman contacto en el estado nativo cuando se encuentran a una distancia que sólo difiera de la distancia en el estado nativo para ese mismo par de residuos en un porcentaje determinado. En nuestro caso es de un 10 % para los contactos de largo alcance y un 5 % para los contactos de corto alcance. De una manera muy similar operamos para el cálculo de los que denominamos **contactos nativos hinchados**. Consideramos que existe un contacto nativo hinchado entre un par de residuos i,j cuando ese par de residuos forman contacto en la estructura nativa y además la distancia entre ese par de residuos ha aumentado en 3 Å con respecto a la distancia que existe para este par en la estructura nativa. Los contactos hinchados nos muestran qué interacciones entre pares de residuos se han desplazado desde el primer pozo de interacción, donde se estabilizan

los contactos nativos, hasta el segundo pozo de interacción de nuestro potencial (Figura 2.7). Para estos contactos nativos también hemos establecido una tolerancia de un 10 % con respecto a la distancia $d_{ij}^{nat} + 3 \text{ \AA}$.

Otras de las propiedades estructurales que obtenemos es el **radio de giro**^{141,142} (R_g). Se trata de una media del valor cuadrático medio de la distancia de los puntos que componen una estructura con respecto al centro de masas de la misma. Se puede calcular a partir de la siguiente ecuación:

$$R_g = \sqrt{\frac{1}{N} \sum_{i=1}^N (r^{ca}(i) - r_{CM})^2} \quad (2.9)$$

donde $r^{ca}(i)$ y r_{CM} son la posición del C α del residuo i y del centro de masas de la cadena, respectivamente. Por tanto el radio de giro muestra una medida de cuán compacta puede ser una configuración determinada. Valores bajos de radio de giro son característicos de estructuras compactas, mientras que valores mayores indican que la configuración es más abierta y menos compacta. Puesto que cada estructura nativa tiene un radio de giro determinado, en este trabajo se presenta siempre R_g^* :

$$R_g^* = \frac{R_g}{R_g^{nativa}} \quad (2.10)$$

como un valor normalizado con respecto al valor del radio de giro de la estructura nativa. De esta manera podemos comparar valores de radio de giro para proteínas de diferente estructura y tamaño. El último parámetro estructural que registramos es el **root mean square deviation**¹⁴³ (RMSD); se trata de una medida de la similitud existente entre dos estructuras determinadas. Para su cálculo se superponen las dos estructuras a comparar, y se calcula la desviación cuadrática media entre centros equivalentes a partir de la siguiente expresión:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_1^{ca}(i) - r_2^{ca}(i))^2} \quad (2.11)$$

siendo r_1^{ca} y r_2^{ca} la posición del C α de cada una de las dos estructuras a comparar. El RMSD se puede calcular para una proteína completa o para un fragmento de proteína. Este último caso es especialmente interesante para estudiar la formación o no de ciertos elementos de estructura secundaria. Un valor de RMSD = 0 indica que las dos estructuras comparadas son exactamente iguales. Cuanto mayor sea el valor de RMSD, mayor serán las diferencias entre las dos configuraciones comparadas¹⁴⁴. En nuestro caso haremos siempre una comparación de las conformaciones que aparecen en el muestreo con la estructura nativa.

A partir de los datos de energía que se obtienen de la simulación, podemos calcular otra serie de propiedades termodinámicas que nos aportan mucha información sobre el sistema. Una propiedad fundamental en el estudio del plegamiento de proteínas es la curva de capacidad calorífica. Si bien experimentalmente se suele medir la capacidad calorífica a presión constante (C_p), en nuestro caso calculamos la **capacidad calorífica a volumen constante** (C_v) que no difiere prácticamente de la C_p en este caso. Por tanto de forma cualitativa las conclusiones que se pueden obtener a partir de C_v son las mismas que las obtenidas a partir de los valores de C_p . La curva de capacidad calorífica es muy importante en el estudio del plegamiento de proteínas, porque la forma de la curva está relacionada con el tipo de tránsito que se da en el proceso de plegamiento, y de ella podemos obtener información relevante del proceso como la temperatura de transición (T_m) o la posible aparición de tránsitos adicionales entre estructuras intermedias estables. La capacidad calorífica es una medida de las fluctuaciones de la energía a temperatura y volumen constante¹⁴⁵, que podemos calcular según la siguiente expresión:

$$C_v^* = \frac{1}{T^{*2}} [\langle E^{2*} \rangle - \langle E^* \rangle^2] \quad (2.12)$$

Un valor alto de C_v indica que existen fluctuaciones considerables en la energía y por tanto que a esa temperatura existe una mayor dispersión de energías entre las distintas configuraciones accesibles. Por esta razón la forma de la curva nos dará mu-

cha información acerca del tipo de tránsito que sufre la proteína. Los asteriscos de la ecuación (2.12) indican que la C_v , temperatura y energía con las que trabajamos se encuentran en unidades reducidas con respecto a $k_b T_{ref}$, siendo k_b la constante de Boltzmann y T_{ref} una temperatura de referencia. De esta manera manejamos la temperatura y la energía como magnitudes reducidas y adimensionales, que definimos según las siguientes expresiones:

$$T^* = \frac{T}{T_{ref}} \quad (2.13)$$

$$E^* = \frac{E}{k_b T_{ref}} \quad (2.14)$$

Con estas relaciones podremos comparar nuestros resultados de simulación con resultados experimentales siempre y cuando tengamos por lo menos una T_{ref} del sistema a estudiar. En cualquier caso, aun no teniendo una T_{ref} para el sistema estudiado, los resultados obtenidos con nuestra simulación mostrarán cómo influyen de forma cualitativa los distintos factores que estudiamos para el plegamiento de proteínas.

Otra de las propiedades que da mucha información acerca del proceso de plegamiento es el estudio del perfil de energía de Helmholtz a las distintas temperaturas. La utilización de la técnica del *parallel tempering*, además de asegurarnos un muestreo muy completo del embudo de plegamiento, nos permite obtener información simultánea del sistema a muchas temperaturas distintas. Si utilizamos la información que obtenemos a cada temperatura y nos ayudamos del método de histogramas ponderados o *weighted histogram analysis method*^{146,147} (**WHAM**), podemos obtener el perfil de energía de Helmholtz a cualquiera de las temperaturas estudiadas. Con los valores de energía y entropía de un estado m , gracias a la ecuación (2.15) podemos calcular la energía de Helmholtz (A_m) a una temperatura determinada,

$$A_m(T) = E_m - T \times S_m \quad (2.15)$$

La entropía del estado m la podemos calcular a partir de la función de densidad de estados Ω_m :

$$S_m = k_b \ln \Omega_m \quad (2.16)$$

La ventaja de este método es que a partir de los datos de la simulación podemos calcular la densidad de cada estado m , dentro de un intervalo de energía alrededor de E_m ¹⁰². Si representamos $A_m(T)$ frente a E_m obtenemos el perfil de energía de Helmholtz a la temperatura T . Estudiando estos perfiles de energía libre podremos conocer las características del proceso de plegamiento a las distintas temperaturas de interés.

Tras haber descrito en este Capítulo detalladamente nuestro método de simulación y nuestro proceso de cálculo, así como el tratamiento numérico que necesitamos para poder estudiar las diferentes propiedades del sistema, podemos pasar a los siguientes capítulos donde analizaremos los resultados obtenidos al aplicar nuestro potencial a diferentes sistemas.

Capítulo 3

Aplicación del potencial de alta presión

Nuestro grupo de investigación lleva años desarrollando distintos modelos basados en la topología del estado nativo para el estudio del plegamiento de proteínas. Sin embargo, aunque se han estudiado diferentes factores que afectan a este proceso y a la metodología utilizada, como la formación de enlaces disulfuro¹⁴⁸ o la influencia del enlace de hidrógeno^{131,132}, siempre se ha realizado el estudio a presión atmosférica. Por tanto, en esta Tesis nos encontramos ante un desafío nuevo, ya que el efecto que puede ejercer la presión en este proceso debe ir más allá de un simple desplazamiento del equilibrio o de una mayor exactitud en los resultados obtenidos. En este caso la transición desde el estado nativo al estado desnaturalizado de la proteínas puede variar notablemente al modificar de forma considerable el embudo de plegamiento que estamos estudiando en cada momento, como esquematizamos en la Figura 2.1.

Los algoritmos que se habían desarrollado en nuestro grupo de investigación estaban suficientemente probados y optimizados para el estudio de la termodinámica del proceso a presión atmosférica. Sin embargo, supone una incógnita conocer y entender cómo un nuevo potencial de interacción, que debe estabilizar a presiones elevadas un estado desnaturalizado de la proteína que difiere notablemente de la conformación *random coil*, afecta a nuestro sistema y sobre todo al proceso de desnaturalización. De hecho, la utilización de la técnica del *parallel tempering* supone una ventaja adicional en este

estudio, no sólo por los beneficios del uso de esta técnica que ya explicamos en el Capítulo anterior, sino porque nos permite estudiar fácilmente el efecto de la presión junto a la temperatura, ya que la simulación la realizaremos a diferentes temperaturas a cada una de las presiones de trabajo. De esta forma será más sencillo explorar el diagrama de fases de las proteínas, ya que estudiaremos nuestro sistema a diferentes presiones y en un rango amplio de temperaturas.

3.1. Motivación del Capítulo

En el Capítulo anterior hemos explicado cada uno de los pasos que seguimos para poder realizar nuestra simulación y cómo obtenemos algunas propiedades termodinámicas a partir de sus resultados. En este Capítulo vamos a mostrar los resultados obtenidos al aplicar nuestro modelo para la simulación de una proteína real. Nuestro objetivo en esta Sección es entender cómo funciona el potencial que hemos desarrollado, y cómo a través de él podemos reproducir el efecto de aplicar una presión hidrostática al sistema que estamos estudiando. Esto es equivalente a encontrar la relación que existe entre los parámetros de nuestro potencial y la presión a la que se encuentra el sistema. Por otro lado, queremos conocer los diferentes mecanismos de plegamiento o desnaturalización que podemos encontrar para una proteína a distintas presiones. Realizando simulaciones en un intervalo amplio de presiones podemos ir observando la aparición de distintos procesos, al mismo tiempo que observaremos cómo se estabilizan distintos estados desnaturalizados de una misma proteína.

Dividiremos el Capítulo en dos partes. En la primera de ellas describiremos con detalle algunos factores que hemos debido tener en cuenta de cara a la implementación definitiva de nuestro modelo de simulación. Para ello nos ayudaremos de los resultados que obtenemos en la simulación de una proteína determinada. Posteriormente analizaremos, tras explicar brevemente nuestras herramientas de análisis, los resultados obtenidos para esta proteína a distintas presiones intentando entender cómo evoluciona el

sistema al aumentar la presión del medio. El objetivo del Capítulo es poder entender perfectamente cómo el potencial actúa sobre el comportamiento del sistema estudiado, y podamos tener una noción de las ventajas e inconvenientes que proporciona este modelo cuando estudiamos proteínas en función de la presión.

3.2. Selección de la proteína

Para la optimización de nuestro potencial hemos utilizado una proteína que ha sido ampliamente estudiada; se trata del dominio B1 de la proteína G de *Streptococcus*, cuyo código PDB¹³⁷ es **2GB1**¹⁴⁹. A lo largo de esta Tesis utilizaremos el código PDB de cada proteína para referirnos a ella. De esta manera hacemos siempre referencia a una estructura determinada para cada proteína, que es la que podemos obtener en esta base de datos y define nuestro mapa de contactos. La selección de esta proteína para analizar nuestro potencial de interacción no es arbitraria, ya que el plegamiento de esta proteína ha sido muy estudiado tanto por medio de simulación^{78,150,151} como por grupos experimentales^{152–154}. Además, nuestro propio grupo de investigación ya ha utilizado esta misma proteína para desarrollar y probar anteriores potenciales y conocemos bastante bien el proceso de plegamiento a presión atmosférica¹⁰². Esta proteína presenta unas características que la convierten en una proteína ideal para poder analizar nuestro potencial:

- Se trata de una proteína compuesta por **una única cadena polipeptídica**. Si se tratase de una proteína compuesta por más de una cadena sería necesario introducir una nueva serie de condiciones en nuestra simulación. En nuestro grupo de investigación ya se abordó el estudio del plegamiento de más de una cadena polipeptídica adaptando el modelo de presión atmosférica para una única cadena¹⁵⁵.
- Se trata de una proteína **sin aminoácidos de prolina en configuración *cis***, algo necesario para que nuestro modelo pueda funcionar adecuadamente. Como se

comentó en la Sección 1.1.1, el enlace peptídico presenta cierto carácter de doble enlace, siendo mayoritaria la configuración *trans*, por su menor impedimento estérico⁷, sin embargo la prolina es un aminoácido donde el impedimento estérico del isómero *cis* no es tan grande, y por tanto estadísticamente es el único aminoácido donde se puede encontrar un enlace peptídico en posición *cis* con el residuo anterior en un porcentaje no despreciable de los enlaces peptídicos que forma¹⁵⁶. Cuando el enlace peptídico se encuentra en posición *cis* y no *trans*, la distancia entre los C α de esos dos aminoácidos es menor a la distancia típica entre carbonos α vecinos en secuencia (3.8 Å). Nuestro modelo necesitaría de pequeñas correcciones para que pudiese ser utilizado para la simulación de proteínas con esta característica.

- La proteína 2GB1 es una proteína que **no necesita de otras moléculas para estabilizar el estado nativo**. Existen muchas proteínas que necesitan de otras moléculas para poder estabilizar el estado nativo y ser biológicamente funcionales¹⁵⁷. Pueden ser pequeños fragmentos de carácter peptídico, o moléculas individuales que actúan como cofactores. Sin embargo, a día de hoy estas proteínas no pueden ser simuladas con nuestro modelo, porque modificarían la forma en la que el mapa de contactos nativos define las interacciones del modelo.
- Esta proteína **no tiene enlaces disulfuro** entre aminoácidos de cisteína. El enlace disulfuro, característico de ciertas proteínas, presenta mayor fortaleza que cualquier otra interacción intramolecular, ya que es sólo ligeramente menor que un enlace covalente¹⁵⁸. Proteínas con este enlace no pueden ser simuladas directamente con este modelo. En caso de querer simular una proteína con dicha característica sería necesario hacer una modificación en el modelo para que este pudiese ser utilizado. Dicha alteración del modelo ya fue realizada por Larriva y col. para el modelo que utilizamos a presión atmosférica¹⁴⁸.
- La proteína 2GB1 es **una proteína de tamaño pequeño**. Para nuestro modelo

el tamaño de la proteína no es un problema directo. Sin embargo, como ya comentamos, un mayor tamaño implica un mayor coste computacional. No olvidemos que la mayoría de las proteínas son macromoléculas con centenares o millares de aminoácidos, y por tanto las pruebas necesarias para poner a punto el programa de simulación resultarían ser enormemente costosas. Por esta razón, una proteína de pequeño tamaño nos permite optimizar tiempo y recursos.

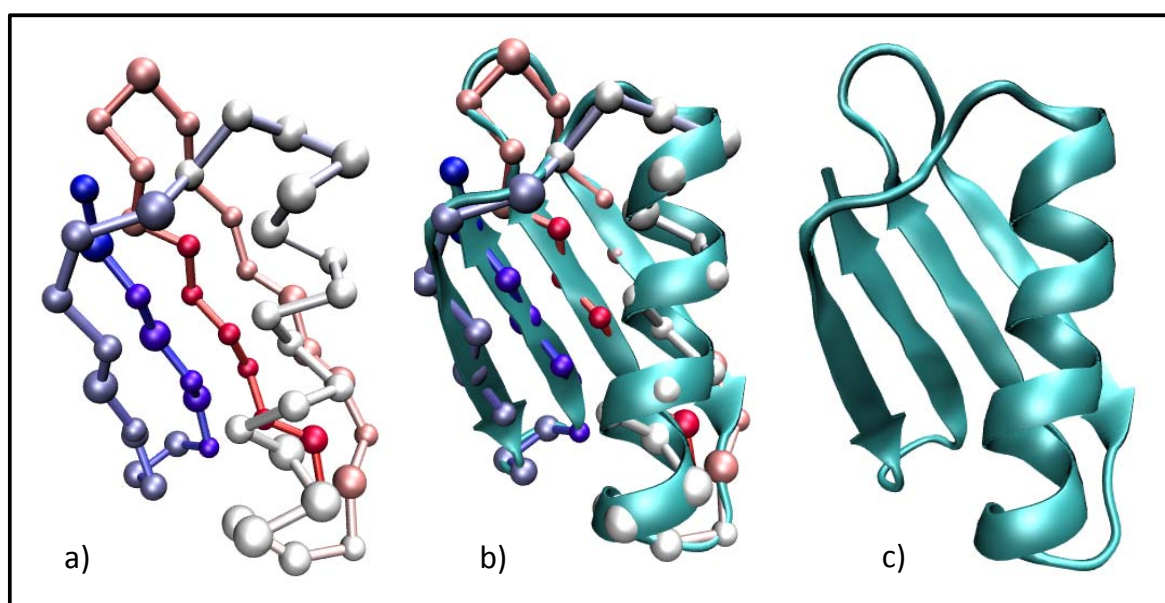


Figura 3.1: **Representaciones de la proteína 2GB1.** a) Representación de $C\alpha$, con una graduación del color a medida que avanzamos en la cadena principal, el primer aminoácido se muestra en rojo y el último en azul, entre medias de color blanco. b) representación simultánea de a) y c). c) Representación esquemática de la estructura secundaria para la misma proteína, con flechas se muestran las hebras β y en espiral se muestra la hélice α .

Las características señaladas son comunes a casi todas las proteínas que se presentarán a lo largo de esta Tesis. Sin embargo, siempre existe la posibilidad de adaptar el modelo con algunas de las características que hemos ido mencionando, de este modo podremos estudiar en el futuro proteínas que tengan una mayor complejidad en su estructura.

La proteína 2GB1, que podemos ver en la Figura 3.1, está compuesta por 56 aminoácidos, los cuales forman una estructura muy simétrica, compuesta por una lámina β enfrente a un hélice α . Se trata de un motivo estructural, lámina β frente a hélice α ,

muy común entre las proteínas de pequeño tamaño¹⁵⁹. Es, por tanto, una proteína de tipo α - β , que nos permitirá estudiar cómo afecta el modelo a los dos tipos de estructura secundaria más comunes dentro de las proteínas, si bien un estudio detallado de esta influencia se tratará en el Capítulo 4.

Todos los resultados que se muestran a lo largo de este Capítulo están referidos a esta proteína.

3.2.1. Mapa de contactos de la proteína 2GB1

Antes de iniciar nuestra simulación es necesario obtener el mapa de contactos de la proteína que vamos a estudiar, ya que la información presente en este mapa es la que determina el potencial de interacción del sistema. En la Figura 3.2 podemos observar el mapa de contactos para la proteína 2GB1. Como sucedía en la Sección 2.2.2, mostramos un mapa de contactos dividido en dos triángulos. Podemos observar cómo para esta proteína tenemos un total de 215 contactos, de los cuales 107 son de corto alcance, 81 son de largo alcance y 27 son contactos de enlace de hidrógeno.

En el triángulo superior mostramos los contactos coloreados en función del tipo de contacto que forman. Como explicamos en la Sección 2.2.2, es necesario diferenciar los tres tipos de contactos con los que trabajamos. Recordemos que sólo los contactos de largo alcance pueden pasar a ser contactos hinchados. Para la selección de los contactos de enlace de hidrógeno utilizamos el programa de simulación *Pymol*¹³³, tal y como explicamos en la Sección 2.2.2. Para esta proteína podemos observar la existencia de cuatro hebras β : Hbr1, Hbr2, Hbr3 y Hbr4, de color amarillo en el esquema de la Figura 3.2. Las cuatro hebras se disponen de una manera bastante especial, ya que forman dos horquillas β : la Hor1 formada por la unión antiparalela² de Hbr1 y Hbr2, y la Hor2 formada por la unión antiparalela de Hbr3 y Hbr4. Estas dos horquillas β a su vez se unen de forma paralela, a través de Hbr1 y Hbr4 formando una lámina β . Precisamente en las zonas de interacción entre hebras β es donde se forman parte de los enlaces

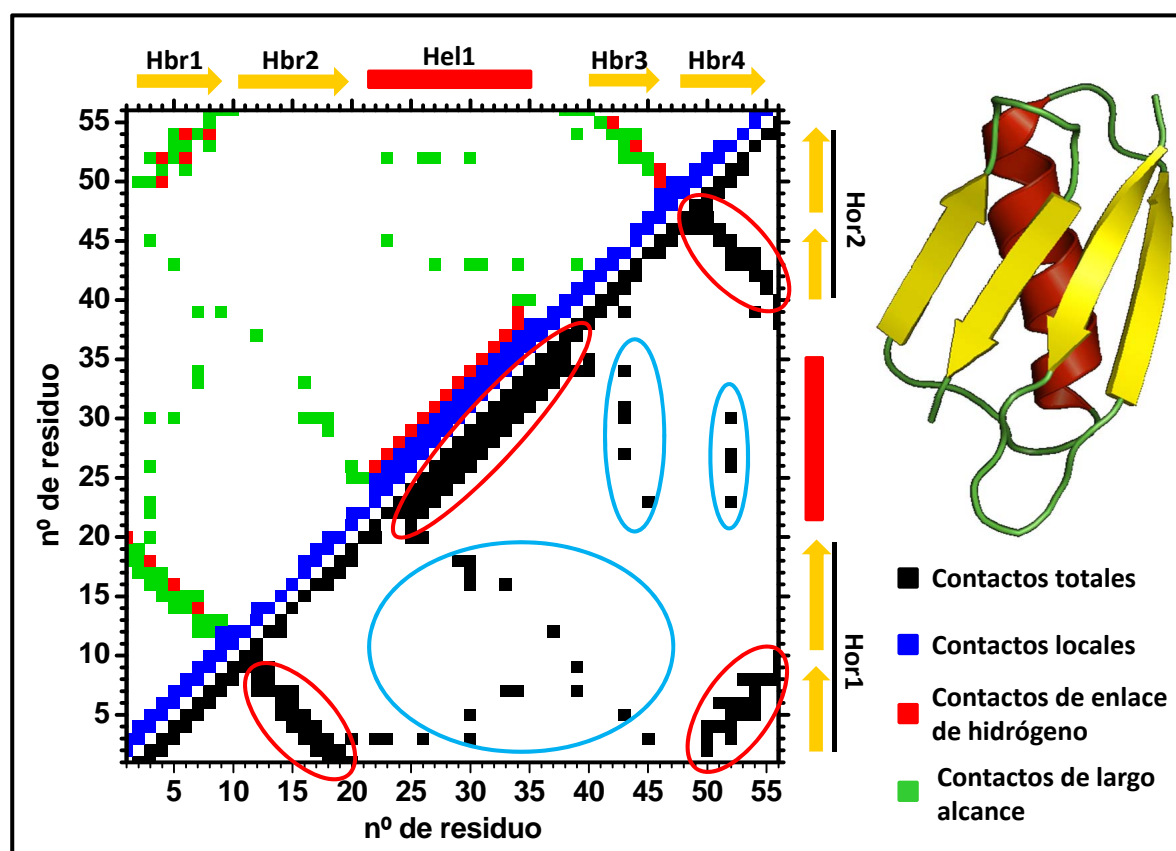


Figura 3.2: Mapa de contactos de la proteína 2GB1. Los contactos se encuentran diferenciados en función del tipo de contacto que forman. Además se muestra una representación esquemática de la proteína 2GB1, para así apreciar fácilmente sus elementos de estructura secundaria.

de hidrógeno que mantienen este tipo de estructuras, como podemos ver en el mapa de contactos. Además, esta proteína tiene una hélice α : Hel1, que se encuentra situada por secuencia entre Hbr2 y Hbr3, y especialmente se sitúa enfrentada a la lámina β . A lo largo de la hélice se pueden observar en el mapa de contactos los diferentes enlaces de hidrógeno que ayudan a su estabilidad.

En el mapa de contactos podemos observar además diferentes contactos de unión entre las diferentes elementos de estructura secundaria que contribuyen a la estabilidad de la estructura nativa. Estos contactos a su vez se pueden dividir en dos tipos: contactos responsables de estabilizar la estructura secundaria y contactos responsables de estabilizar la estructura terciaria de la proteína. Los primeros son los que se forman entre aminoácidos que pertenecen al mismo elemento de estructura secundaria, Estos son,

junto con los contactos de enlace de hidrógeno, los responsables de la estabilidad de las dos horquillas β , de la lámina β y de la hélice α . En el triángulo inferior de la Figura 3.2 se señalan estos contactos con elipses de color rojo. El segundo tipo de contactos, señalados en la Figura con elipses de color azul, son los que se producen entre aminoácidos de diferentes elementos de estructura secundaria que se encuentran espacialmente cercanos en la estructura nativa. Son contactos necesarios para que la proteína mantenga una estructura compacta y globular, y por tanto son los responsables de mantener la estructura terciaria de la proteína.

3.3. Calibrado del potencial de interacción

Una vez seleccionada la proteína, el tipo de simulación que vamos a desarrollar y el potencial de interacción que vamos a aplicar, ya estamos en disposición de empezar a realizar nuestros **experimentos virtuales**. Sin embargo, sigue quedando una duda sin resolver y es: ¿cómo podemos reproducir el efecto de aplicar una presión hidrostática y creciente al sistema a través del potencial de interacción que hemos desarrollado? Nuestro potencial está compuesto por dos pozos de interacción y una barrera de desolvatación, y aunque existe una relación entre la profundidad de dichos pozos y barrera con la presión del sistema, expresión (2.3), sin embargo dicha correspondencia no está todavía cuantificada. Por tanto, para poder calibrar nuestro potencial es necesario encontrar una relación entre la presión del sistema y el valor de la relación de los diferentes parámetros $\epsilon_{db}/\epsilon_{go}$ y $\epsilon_{ssm}/\epsilon_{go}$. En el trabajo de Garde y col.¹²⁶ que se muestra en la Figura 2.6 (b) y otros trabajos similares^{127,128,160}, se muestran valores determinados de estos tres parámetros, que corresponden con una presión determinada. Existe una relación cuantitativa entre el valor que toman los distintos ϵ y la presión del sistema, para la interacción entre dos grupos hidrófobos.

Con estos antecedentes, en una primera etapa de esta Tesis realizamos un barrido completo de valores de ϵ para los tres parámetros. De esta manera abordamos un estudio

sistemático de la relación que tienen entre sí los tres parámetros, intentando encontrar una relación cuantitativa entre los valores de los distintos ϵ y la presión del sistema. Procedimos a realizar una batería de simulaciones donde hicimos variar todos los valores de ϵ_{go} , ϵ_{db} y ϵ_{ssm} en un rango entre 0 y 1. Esto supuso un número considerable de simulaciones, puesto que para estudiar la evolución de cada parámetro es necesario realizar diferentes simulaciones manteniendo uno de los parámetros constante y hacer las distintas combinaciones posibles de los otros dos. Sin embargo, al estudiar los resultados de las distintas simulaciones nos encontramos con que no obtuvimos una relación directa entre los valores que adoptaban los tres parámetros y los resultados esperables para una presión característica en nuestro sistema. Posiblemente la relación sea mucho más complicada y se necesite de una combinación matemática compleja de los tres parámetros para poder encontrar la relación directa entre ellos y la presión que queremos reproducir en el sistema. Además, esta metodología nos mostró un problema importante: tal y como está planteado el modelo, si el valor de ϵ_{go} es variable, y puesto que es este el valor sobre el que está normalizada la energía, como vimos en la Sección 2.2.1, la energía deja de ser una propiedad idónea para comparar las diferentes configuraciones. En este caso, configuraciones procedentes de simulaciones donde el valor de ϵ_{go} sea distinto tendrán siempre diferente energía pese a ser muy similares entre sí, algo que afectaría incluso al estado nativo. Esto supone un pequeño inconveniente, ya que el valor de energía en un modelo tipo $G\bar{o}$ representa una buena propiedad para poder comparar configuraciones bajo distintas condiciones. De hecho, esta es la propiedad que se venía utilizando en el grupo para identificar las distintas configuraciones que se obtienen a lo largo de la simulación.

Por tanto nos planteamos otra vía, a la postre definitiva, para poder ‘calibrar’ nuestro potencial. Decidimos que el valor de ϵ_{go} debía permanecer constante para cualquier presión. De esta manera conseguimos que nuestra escala de energía no varíe de forma tan drástica en función de la presión a la que nos encontremos. Si el valor de ϵ_{go}

es constante a cualquier presión la unidad de energía tendrá siempre el mismo valor, por esa razón decidimos utilizar un valor de 1.0 como el valor constante de ϵ_{go} . De esta manera conseguimos que al menos la conformación nativa a presiones diferentes tenga siempre el mismo valor de energía. Además, para simplificar notablemente el potencial y puesto que la parte repulsiva (la barrera de solvatación) de nuestro potencial resulta ser menos influyente que la atractiva, decidimos mantener constante la relación entre el parámetro ϵ_{go} y el parámetro ϵ_{db} . En base a otros trabajos donde se realizó un tratamiento similar^{100,125}, al parámetro ϵ_{db} le asignamos un valor de un 30 % del valor de ϵ_{go} , por tanto ϵ_{ssm} lo mantendremos constante e igual a 0.3:

$$\epsilon_{go} = 1,0 ; \frac{\epsilon_{db}}{\epsilon_{go}} = 0,3 \quad (3.1)$$

De esta forma conseguimos un potencial de interacción fácilmente adaptable a diferentes presiones, puesto que sólo depende de un único parámetro, ϵ_{ssm} , reduciendo de tres a tan sólo uno los parámetros de nuestro potencial que reproducen dentro de una aproximación de campo medio, el efecto de la presión.

$$\text{Presión de sistema} \propto \frac{\epsilon_{ssm}}{\epsilon_{go}} \quad (3.2)$$

De esta forma, en nuestro modelo simplificado, un aumento del pozo ssm implica un aumento de la presión del sistema, como también se aprecia en los resultados bibliográficos de la Figura 2.6 (b). Así, un **valor alto de ϵ_{ssm} implica que nos encontramos a presión elevada**, mientras que un **valor nulo o muy pequeño de ϵ_{ssm} es indicativo de que el sistema se encuentra a presión atmosférica**. Se trata en este momento de una relación cualitativa y no cuantitativa, pero es una relación sencilla, que nos permitirá estudiar un sistema a presiones crecientes, cuando realicemos simulaciones donde el valor de ϵ_{ssm} vaya aumentando. Precisamente esa será la forma de trabajar con cada sistema, realizaremos simulaciones con diferentes valores de ϵ_{ssm} que corresponderán al plegamiento a distintas presiones. No conocemos la presión exacta del sistema en ningún momento, pero sí sabemos de forma cualitativa cuándo una presión

es mayor que otra.

En la siguiente Sección mostraremos los resultados obtenidos al realizar la simulación del plegamiento de la proteína 2GB1 a distintas presiones. Al ser los primeros resultados que mostramos, prestaremos especial interés al tratamiento numérico que debemos realizar para transformar las diferentes configuraciones obtenidas de la simulación en información sobre el plegamiento de una proteína concreta.

3.4. Herramientas para el estudio del tipo de plegamiento

3.4.1. La curva de capacidad calorífica

Como ya comentamos en la Sección 2.4, la curva de capacidad calorífica (C_v) resulta ser determinante a la hora de estudiar nuestro proceso de plegamiento^{161,162}, ya que de su forma podremos obtener información valiosa acerca del tránsito estudiado. A continuación, señalaremos las características más importantes del comportamiento de C_v , que nos proporcionan información acerca de los diferentes estados que puedan estar involucrados en el tránsito entre el estado nativo y el estado desnaturalizado.

- **La temperatura del máximo de la curva.** La temperatura a la que encontramos el máximo de C_v corresponde a la temperatura a la que se produce la máxima amplitud en las fluctuaciones del sistema, e indicará normalmente la temperatura central del tránsito entre el estado nativo de la proteína y su estado desnaturalizado. La denominamos temperatura de transición (T_m). Puesto que la curva de capacidad calorífica muestra las fluctuaciones de energía que existen entre el total de conformaciones a cada temperatura, el máximo en la curva nos muestra la temperatura a la cual existe mayor fluctuación entre todos sus estados. En el caso de un tránsito de dos estados esto ocurrirá justo a la temperatura de equilibrio donde

aproximadamente un 50 % de las conformaciones corresponden al estado nativo y un 50 % corresponden al estado desnaturalizado. La temperatura de transición es característica de cada proteína y se puede medir experimentalmente. Precisamente, la medida experimental de T_m puede servir como temperatura de referencia de cara a comparar nuestros resultados con los resultados experimentales. A partir del valor de T_m experimental y gracias a la ecuación (2.13), podemos obtener temperaturas y energías reales del sistema a partir de nuestros valores de temperaturas y energías reducidas.

- **La forma de la curva.** La altura y anchura de la curva están relacionadas entre sí, puesto que la integral bajo la curva de capacidad calorífica esta relacionada con la diferencia de energía entre el estado de mayor temperatura y el estado de menor temperatura. Para una misma proteína, la integral bajo la curva tiene entonces que permanecer constante si los estados de partida y de llegada son los mismos, o al menos suficientemente parecidos. Por tanto, para la misma proteína en las mismas condiciones, curvas muy estrechas serán siempre muy altas, mientras que curvas muy anchas serán siempre muy bajas. La forma de la curva está relacionada con el tipo de tránsito: **curvas muy altas** son indicativas de que existen muchas diferencias entre estados que son estables a la misma temperatura, lo que **es característico de plegamientos de dos estados**, en donde dos estados bien diferenciados se encuentran en equilibrio, y por tanto las fluctuaciones de energía del sistema son muy amplias a la temperatura de equilibrio entre los diferentes estados involucrados. En la Figura 3.3 (a) la curva de capacidad calorífica de color rojo muestra un ejemplo de este tipo de tránsito. Por otro lado **curvas** de capacidad calorífica **muy anchas** son características de tránsitos continuos tipo downhill donde el tránsito entre un estado y otro es paulatino a través de configuraciones intermedias, con una población que va evolucionando desde un estado a otro de forma continua. La curva de color negro en la Figura 3.3 (a) es un ejemplo

de este tipo de tránsito. Relacionado con el tipo de tránsito existe un concepto importante de cara a nuestras simulaciones que no debemos olvidar. En los tránsitos muy cooperativos, por tanto muy estrechos en la escala de temperaturas, el equilibrio se da en un intervalo muy pequeño de temperaturas. Esto hace que para poder realizar un muestreo adecuado por *parallel tempering* sea necesario hacer simulaciones a temperaturas muy cercanas, para que los histogramas de energía entre temperaturas contiguas puedan solapar correctamente, como comentamos en la Sección 2.3 al discutir el procedimiento. Esto conlleva introducir muchas temperaturas en la simulación, y por tanto un mayor coste computacional. En general, esto hace que con nuestro modelo equilibrios de dos estados excesivamente cooperativos se hagan más difíciles de estudiar, no por una mayor complicación del sistema, sino por un aumento drástico del tiempo de simulación, ya que se necesitan muchísimas temperaturas a lo largo de la transición para simular correctamente tránsitos muy cooperativos. Este problema no existe en tránsitos poco cooperativos, donde hace falta barrer un intervalo algo mayor de temperaturas, pero al producirse un tránsito gradual las temperaturas a las que realizamos las simulaciones pueden estar más separadas y no son necesarias tantas réplicas en paralelo para muestrear todo el embudo de plegamiento. Además, esta situación agrava el problema de la simulación de proteínas con muchos aminoácidos, ya que cuanto mayor es el tamaño de una proteína, más estrecha es su curva en caso de ser una proteína que pliega por un mecanismo de dos estados. Por tanto, el coste computacional de esa simulación no sólo es mayor por el efecto directo del mayor número de aminoácidos, sino que también se hace mayor por el efecto indirecto de tener que utilizar un mayor número de temperaturas en la simulación.

- **Otros máximos en la curva.** La aparición de más de un pico en la curva (curva de color azul en la Figura 3.3 (b)) o de hombros dentro del pico principal (curva de color morado en la misma Figura) es característica de tránsitos con más de

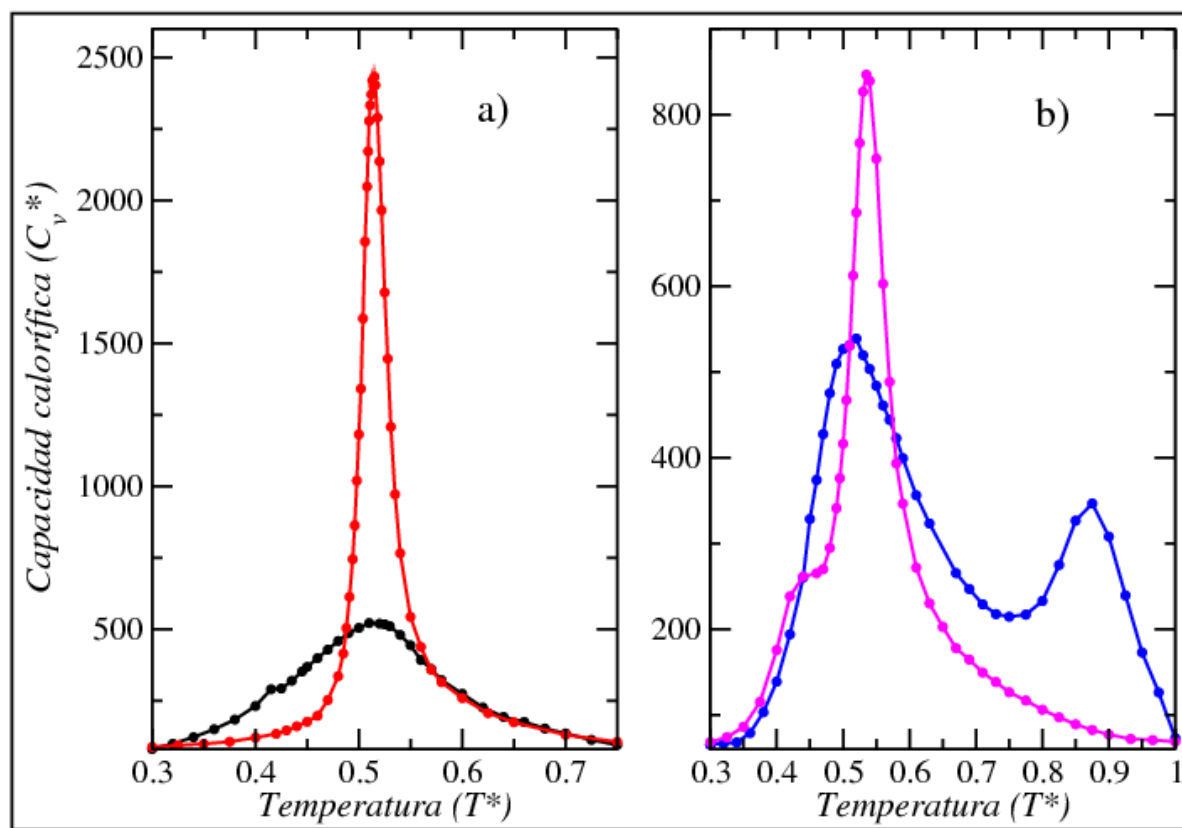


Figura 3.3: **Ejemplos de curvas de capacidad calorífica.** a) Comparación de dos curvas de capacidad calorífica de diferente anchura. b) Se muestran dos curvas donde aparece más de un pico.

dos estados, ya que cada pico de la curva representa el tránsito entre dos estados diferentes con poblaciones apreciables. La curva debe presentar $n-1$ picos, siendo n el número de estados. Sin embargo, alguno de estos picos podría camuflarse en forma de hombro o de tránsito muy solapado dentro del perfil de un pico mayor, y por tanto algunos de ellos pueden no llegar a ser apreciables. Cuando existen tránsitos de más de dos estados y observaremos más de un pico en la curva de capacidad calorífica, la forma, el tamaño y la T_m de cada uno de los picos que aparecen están relacionados con los tránsitos que haya entre todos los estados que puedan coexistir a la temperatura de cada pico.

De esta manera queda explicado cómo la C_v resulta ser una herramienta esencial para el estudio del proceso de plegamiento de proteínas, ya que con un simple vistazo a la curva de capacidad calorífica se puede obtener muchísima información acerca del

número y tipo de tránsitos para una proteína determinada. Puesto que para cada una de nuestras condiciones de presión realizamos distintas simulaciones independientes, las curvas de capacidad calorífica que presentamos en esta Tesis serán siempre una media de las que obtuvimos para los distintos casos, y las barras de error que se muestran en las gráficas corresponden al error estadístico de las medias, en cada punto.

3.4.2. El perfil de energía de Helmholtz

Otra de nuestras herramientas esenciales para la caracterización del tipo de tránsito en el plegamiento de proteínas son los perfiles de energía de Helmholtz que calculamos gracias al método WHAM¹⁴⁷. Los perfiles de energía de Helmholtz nos muestran, a una temperatura determinada, la energía de Helmholtz de los estados poblados a esa temperatura. Esto resulta muy interesante por dos aspectos principales. Por un lado nos permite estudiar la evolución de la energía de Helmholtz de los diferentes estados presentes en el sistema en función de la temperatura. En la Figura 3.4 observamos el perfil de energía obtenido de nuestras simulaciones a diferentes temperaturas para la proteína 2GB1 a presión atmosférica, utilizando un valor de ϵ_{ssm} igual a 0.00. Se observa cómo a bajas temperaturas se estabiliza un estado de baja energía que corresponde al estado nativo de la proteína (N), en la zona izquierda de la representación. A medida que aumenta la temperatura, el perfil de energía va evolucionando para dar lugar a la aparición de un segundo máximo de mayor energía, que corresponde al estado desnaturalizado (D), que aparece en la zona derecha de la representación. En la Figura se puede apreciar cómo la evolución es paulatina, y cómo por debajo de T_m , que para este valor de ϵ_{ssm} es de 0.601 y corresponde a la curva de color negro, es mucho más estable el estado nativo, mientras que por encima de dicha temperatura se estabiliza el estado desnaturalizado, siendo precisamente T_m la que dictamina la frontera de estabilidad entre ambos estados.

Otro aspecto interesante en estas representaciones está en el perfil de energía a la T_m correspondiente. Podemos observar cómo para esta temperatura ambos estados son

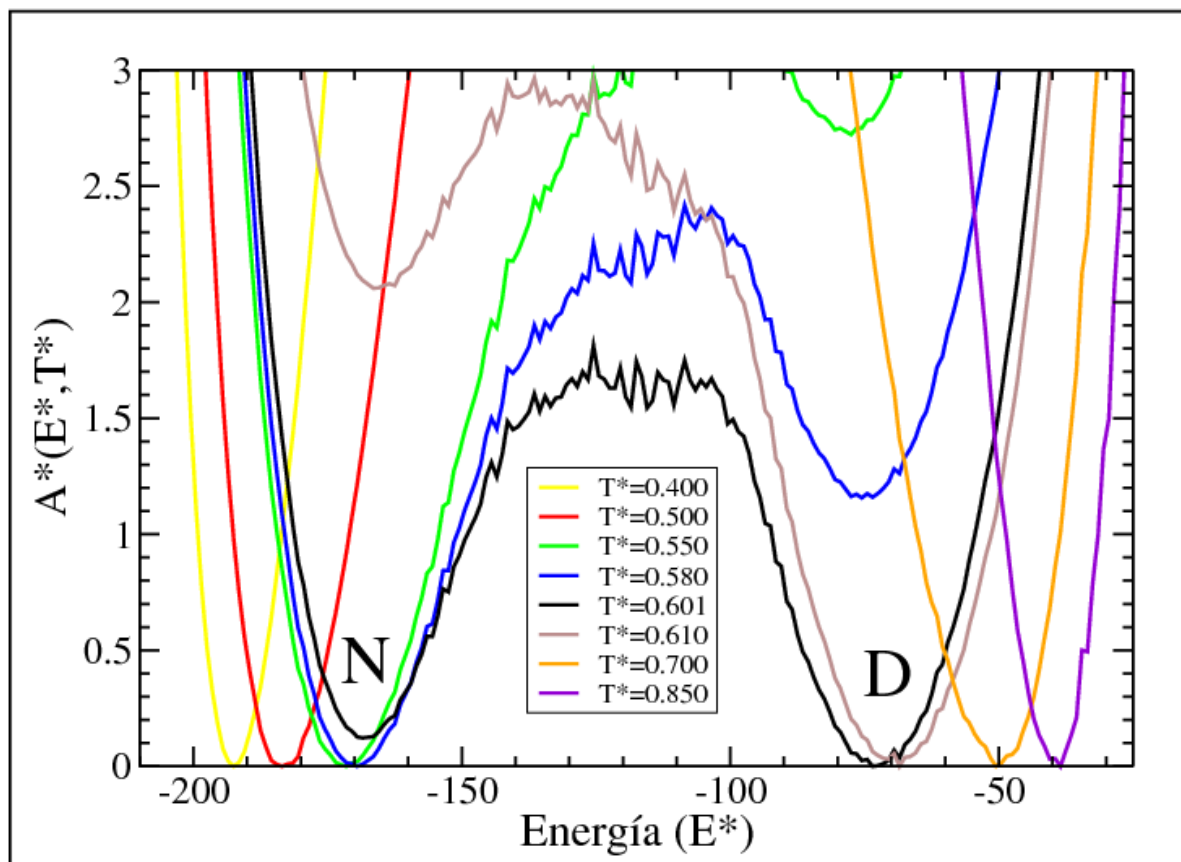


Figura 3.4: **Perfil de energía de Helmholtz para 2GB1 a presión atmosférica.** Evolución de los distintos perfiles de energía de Helmholtz en función de la temperatura. Sólo a T_m ($T^* = 0.601$) son igualmente estables los estados nativo y desnaturalizado.

estables y coexisten, con una barrera energética entre ambos, lo cual es característico de un plegamiento cooperativo de dos estados, como es el que se describe en la bibliografía para esta proteína a presión atmosférica^{163,164}. Precisamente los perfiles de energía de Helmholtz resultan ideales para poder observar cómo varía esta barrera energética entre ambos estados, ya que en caso de ser un tránsito no cooperativo, dicha barrera sería pequeña o inexistente. Por tanto, la representación de los perfiles a la temperatura de transición de la proteína resulta ser especialmente reveladora para conocer ante qué tipo de tránsito nos encontramos, confirmando los resultados de las curvas de capacidad calorífica. Otro aspecto que podemos observar en esta Figura es **el efecto de la temperatura en el estado nativo**. Como se puede observar, la energía del estado nativo de la proteína, pese a ser ‘único’ para cada proteína, evoluciona sensiblemente con la

temperatura. Es claramente visible cómo a la temperatura más baja, $T^* = 0.4$, el estado nativo está definido en un valor de energía de alrededor de -190 unidades reducidas (u.r.). Sin embargo a T_m , el estado nativo se define alrededor de -170 u.r. Esta diferencia se debe a las **fluctuaciones térmicas del sistema**, y es que al ser la temperatura cada vez mayor, la energía del sistema también lo es, y por tanto al sistema le es permitido adoptar algunas configuraciones de mayor energía, que representan fluctuaciones estructurales alrededor del estado nativo.

3.4.3. Histogramas y representaciones bidimensionales

La última de las herramientas que utilizamos para analizar el tipo de tránsito ante el que nos encontramos es la representación de alguna de las propiedades que obtenemos de cada una de las configuraciones estudiadas. Puesto que el muestreo que realizamos es bastante amplio, la representación del histograma de cualquiera de las propiedades medidas nos muestra qué valores son mayoritarios para esa propiedad y a partir de esos datos podemos estudiar las conformaciones que representan estos valores. Del mismo modo podremos observar si existe más de un estado con población apreciable y si existen zonas de baja población entre los diferentes máximos, que nos indicarán de forma indirecta la presencia de barreras energéticas. Como ejemplo mostramos la Figura 3.5 (a), donde mostramos los histogramas de energía (E^*), obtenidos para la simulación a presión atmosférica ($\epsilon_{ssm} = 0.00$) a tres temperaturas diferentes. Se pueden observar diferentes estados en función de la temperatura, y cómo a la temperatura de transición coexisten dos estados con una barrera energética entre ambos, como muestra la falta de población que hay en los valores de E^* situados entre los dos estados poblados.

Además, siguiendo esta forma de estudiar las configuraciones, podemos ir un poco más allá realizando representaciones simultáneas de dos de las propiedades estudiadas. De esta manera obtenemos una representación bidimensional. En cada eje presentamos una de las propiedades seleccionadas, y cada punto de la gráfica representa una de las

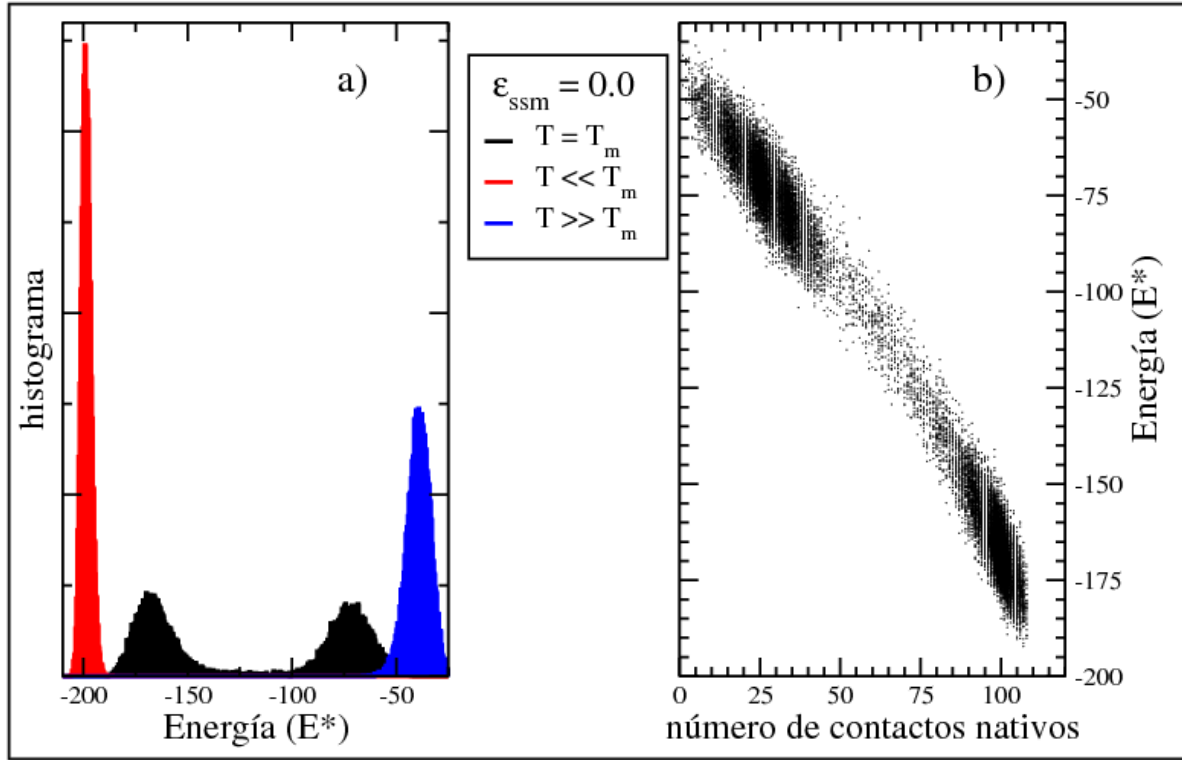


Figura 3.5: **Histograma y representación bidimensional a presión atmosférica.** Se muestran resultados obtenidos a partir de diferentes configuraciones para la simulación a presión atmosférica, $\epsilon_{ssm} = 0.0$. a) Histograma de energía a tres temperaturas diferentes, muy por debajo de T_m , a T_m ($T^* = 0.601$), y muy por encima de T_m . b) Representación bidimensional de la energía frente al número de contactos nativos a T_m .

conformaciones registradas a lo largo de la simulación. Así podemos estudiar en qué zonas de nuestra representación se acumulan un mayor número de conformaciones, obteniendo información tanto del número de estados involucrados (cada estado mostrará un máximo de densidad de puntos en la gráfica) como de la posible presencia de barreras energéticas en caso de no existir población entre las diferentes nubes de puntos. La Figura 3.5 (b) muestra una representación bidimensional de la energía (E^*) frente al número de contactos nativos a la temperatura de transición, siempre a presión atmosférica. Se observa de nuevo la presencia de dos zonas de máxima población, correspondientes a dos estados distintos, y una zona de menor población entre ambas. En esta imagen queda patente una de las carencias de este tipo de representación, la gran cantidad de configuraciones que generamos impide en muchos casos hacerse una idea cuantitativa de la población que existe de cada una de las nubes de puntos. En el ejemplo que mostramos

en la Figura 3.5, cuando nos encontramos a la temperatura de transición, en el histograma es fácil de observar que entre los valores altos y bajos de energía existe una zona de población mucho menor, entre $E^* = -90$ y -140 u.r., aproximadamente. Sin embargo esto puede resultar más difícil de apreciar en la representación bidimensional, ya que si observamos rápidamente la representación, en un primer momento puede parecer que no exista tal barrera. Solamente cuando analizamos con detalle la gráfica, se observa que aunque exista cierta población en la zona intermedia de la representación, la densidad de puntos es, en cualquier caso, muy inferior a la densidad correspondiente a los dos máximos, y por tanto la barrera sí está presente. Esto nos hace ser conscientes de que estos dos tipos de diagramas que utilizamos para el análisis de configuraciones es mejor analizarlos de manera conjunta para así obtener una información más correcta del sistema. En su defecto, podemos usar técnicas de análisis numérico para evaluar el número de casos que aparecen en cada una de las regiones de estas gráficas bidimensionales, como veremos más adelante.

3.5. El tránsito para la proteína 2GB1

3.5.1. Análisis del incremento de presión

En las simulaciones realizadas para el estudio de 2GB1 utilizamos un valor diferente y creciente de ϵ_{ssm} , de tal forma que a medida que aumenta dicho parámetro el sistema se encuentra a mayores presiones, tal y como describimos en la Sección 3.3. El valor de ϵ_{ssm} lo hicimos variar en un rango que va desde 0.00, lo que en nuestro modelo corresponde a la presión atmosférica, hasta un valor de 1.40, que como luego comentaremos corresponde a altas presiones. Entre estos dos valores realizamos simulaciones con los siguientes valores de ϵ_{ssm} : 0.25, 0.50, 0.75, 1.00, 1.10, 1.20 y 1.30. En la Figura 3.6 se muestran las diferentes curvas de capacidad calorífica que hemos obtenido para cada una de las condiciones simuladas. Analizando estas curvas según los criterios establecidos en la

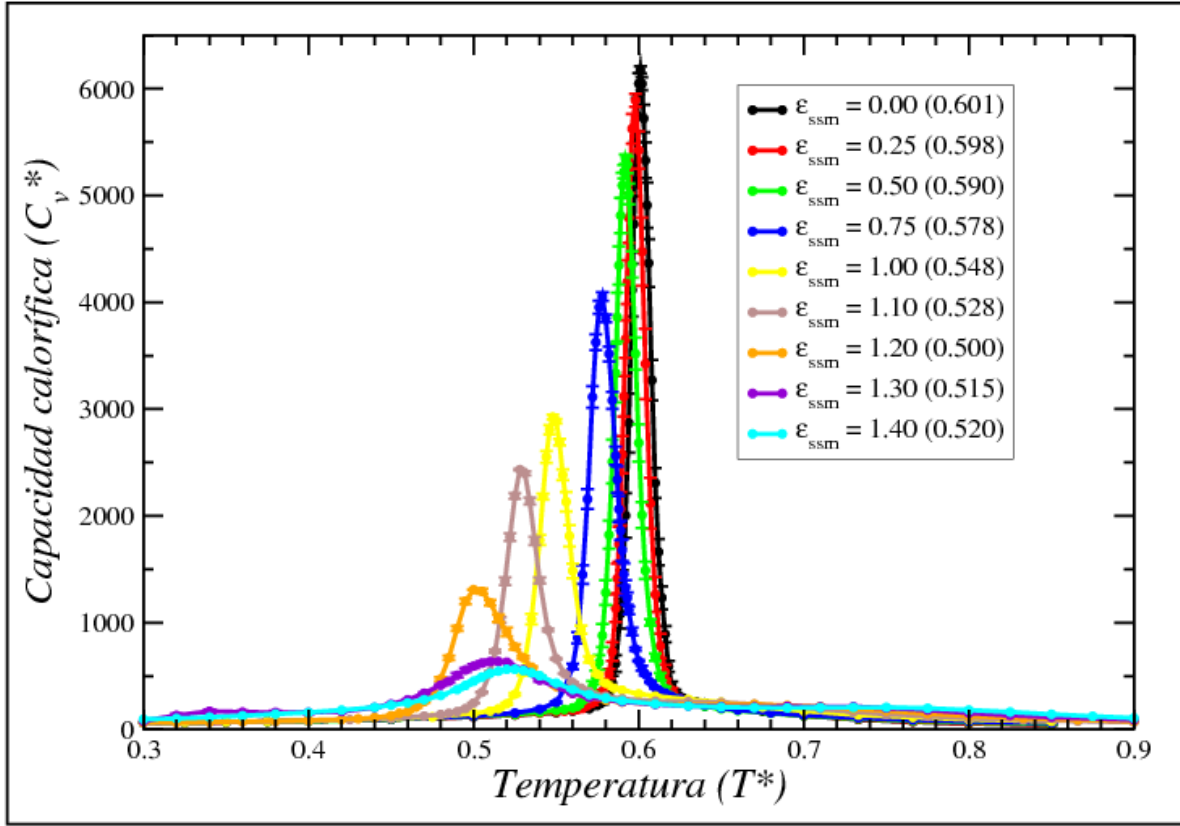


Figura 3.6: **Curvas de capacidad calorífica, en función de la presión.** Mostramos las diferentes curvas de capacidad calorífica que se han obtenido en las distintas simulaciones para valores crecientes de ϵ_{ssm} , en un rango que va desde 0.00 hasta 1.40. En la leyenda se muestra, entre paréntesis, la temperatura T^* del máximo de cada curva, lo que corresponde a su T_m .

Sección anterior, hemos podido observar diferentes aspectos relativos al cambio de las características de la transición a medida que aumenta la presión.

Por un lado, existe un claro desplazamiento de la T_m , que evoluciona normalmente hacia valores menores a medida que aumenta el valor de ϵ_{ssm} . La leyenda muestra entre paréntesis los valores de T_m que hemos obtenido para cada simulación. Esto nos indica que a medida que aumentamos la presión al sistema, la temperatura de desnaturalización va disminuyendo. Este comportamiento sólo se rompe cuando el valor de ϵ_{ssm} es superior a 1.2. Se observa cómo para valores pequeños de ϵ_{ssm} el desplazamiento de T_m es muy pequeño con relación a su valor a presión atmosférica. Sin embargo, cuando el valor de ϵ_{ssm} se aproxima al valor de ϵ_{go} (recordemos que para todos nuestros experimentos ϵ_{go} es igual a 1), el efecto se hace más pronunciado y los desplazamientos son mayores. Cuando

ϵ_{ssm} se hace mayor que 1.2 este comportamiento se invierte, y T_m empieza a aumentar a medida que sigue aumentando la presión del sistema.

$$\begin{array}{ll} \text{Para } \epsilon_{ssm} \leq 1.2 & \uparrow \epsilon_{ssm} = \uparrow \text{presión} \propto \downarrow T_m \\ \text{Para } \epsilon_{ssm} > 1.2 & \uparrow \epsilon_{ssm} = \uparrow \text{presión} \propto \uparrow T_m \end{array}$$

Además, se observa un cambio muy importante en la forma de las curvas. Podemos ver cómo para los valores más bajos de ϵ_{ssm} la curva es alta y estrecha, lo cual es característico de tránsitos muy cooperativos. Sin embargo, a medida que aumenta la presión del sistema lo que observamos es que la curva va adoptando una forma mucho más achatada. Esto nos indica que existe un cambio en el tipo de tránsito, el proceso se va transformando en un tránsito cada vez menos cooperativo.

Si observamos con detenimiento la Figura 3.6, podemos apreciar cómo para las curvas donde el valor de ϵ_{ssm} es igual o superior a 1.00 existe un desdoblamiento del pico principal en dos picos diferentes: uno principal que es fácilmente visible en la Figura 3.6, y uno mucho mas pequeño, a mayores valores de T^* , que se va separando más y más del principal a medida que aumenta la presión del sistema. Para poder observar este fenómeno con más claridad, en la Figura 3.7 mostramos una ampliación de la Figura 3.6 en el intervalo de temperaturas donde aparece el pico secundario. En la ampliación se observa claramente el desdoblamiento del pico principal en los dos picos ya mencionados. En esta imagen es más fácil observar cómo este comportamiento sólo aparece cuando ϵ_{ssm} es igual o mayor a 1.00. Este valor de ϵ_{ssm} es un valor importante en nuestro modelo, ya que coincide con la frontera donde el pozo ϵ_{ssm} empieza a ser tan profundo como el pozo ϵ_{go} , y por tanto, el sistema empieza a estabilizar considerablemente los contactos hinchados con respecto a los contactos directos, tal como los definimos en la sección 2.2.1. Es un desdoblamiento que, para los valores de ϵ_{ssm} iguales a 1.00 y 1.10, supone únicamente un pequeño hombro en el pico principal de la curva de capacidad calorífica, mientras que para valores de ϵ_{ssm} igual o superior a 1.20 supone un pico independiente en la curva. En la leyenda de la Figura 3.7, entre paréntesis, se muestran las temperaturas de

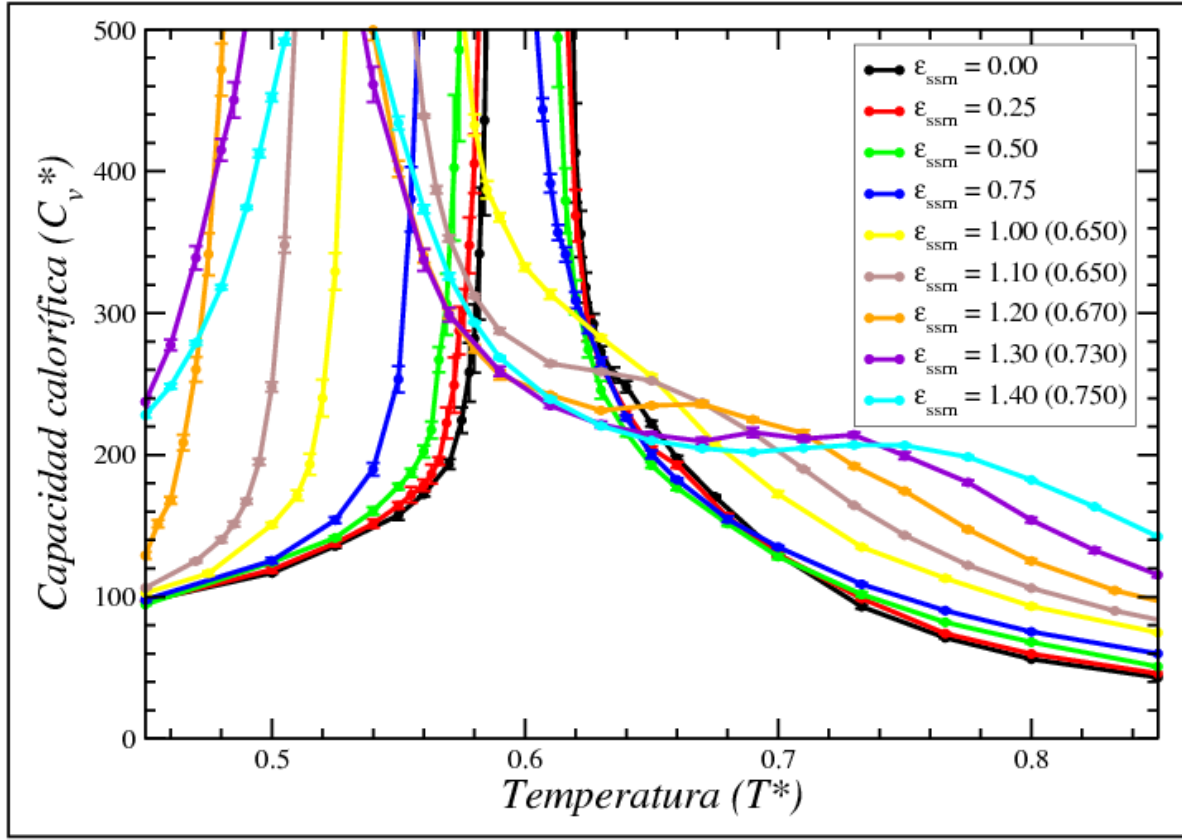


Figura 3.7: **Ampliación de la Figura 3.6, en la zona del pico secundario.** Se muestra una ampliación de la Figura anterior, seleccionando un área de especial relevancia donde se observa la aparición de un segundo pico en las curvas de capacidad calorífica. En la leyenda se muestran entre paréntesis las temperaturas de transición del segundo máximo para aquellas curvas que lo presentan, todas ellas con $\epsilon_{ssm} \geq 1.00$.

los máximos de este segundo pico. La aparición del segundo máximo indica la existencia de por lo menos un tercer estado a lo largo del proceso de desnaturalización. Este tercer estado debe corresponder a un estado intermedio entre el estado nativo y el estado desnaturalizado. Este nuevo estado intermedio parece estabilizarse a medida que aumenta la presión, ya que el segundo máximo de la curva de capacidad calorífica no sólo aumenta en su temperatura de transición, sino que tiende a perfilarse a medida que aumenta la presión del sistema.

Por tanto, a través de un sencillo análisis de la curva de capacidad calorífica, podemos hacernos una primera idea aproximada del efecto de la presión en el proceso de plegamiento, obtenido con nuestro modelo.

- La temperatura de transición, entendida como el máximo absoluto de la curva de

capacidad calorífica, disminuye a medida que aumenta la presión. Esta característica revierte cuando la presión es muy elevada.

- Existe una evolución del tipo de tránsito a medida que aumenta la presión del sistema, pasando de un tránsito cooperativo a un tránsito no cooperativo.
- Con el aumento de la presión notamos la aparición de un nuevo estado.

Con la intención de profundizar en el estudio de los diferentes tipos de transitos que obtenemos al aumentar la presión, procedemos al estudio de los perfiles de energía de Helmholtz a las temperaturas del máximo de las curvas de capacidad calorífica. En la Figura 3.8 mostramos los diferentes perfiles de energía de Helmholtz que hemos obtenido para cada una de las simulaciones a la T_m correspondiente. Un análisis detallado de esta Figura nos permite obtener algunas conclusiones más precisas sobre la evolución del tránsito a medida que aumenta la presión:

- Podemos observar cómo el estado nativo, el de menor energía, permanece prácticamente inalterado hasta que la presión del sistema es suficientemente alta, $\epsilon_{ssm} \geq 1.00$. El pequeño desplazamiento hacia menores energías que se observa al aumentar la presión se debe a las menores fluctuaciones térmicas que sufre el estado nativo, ya que al aumentar la presión nos encontramos ante menores temperaturas de transición. Es importante destacar cómo para valores de ϵ_{ssm} superiores a 1.20, el máximo que observamos en la curva de capacidad calorífica no corresponde a un equilibrio entre el estado nativo y el estado desnaturalizado, como venía siendo habitual hasta esta presión. En estos casos, los valores de energía que obtenemos a la temperatura de transición no corresponden a la presencia de un estado nativo en equilibrio con un estado desnaturalizado, sino que se aprecia la presencia de un único estado con energía similar a la energía del estado desnaturalizado cuando $\epsilon_{ssm} = 1.20$. Este tipo de perfil de energía de Helmholtz es característico de los transitos no cooperativos tipo *downhill*. Por tanto, parece que a altas presiones lo

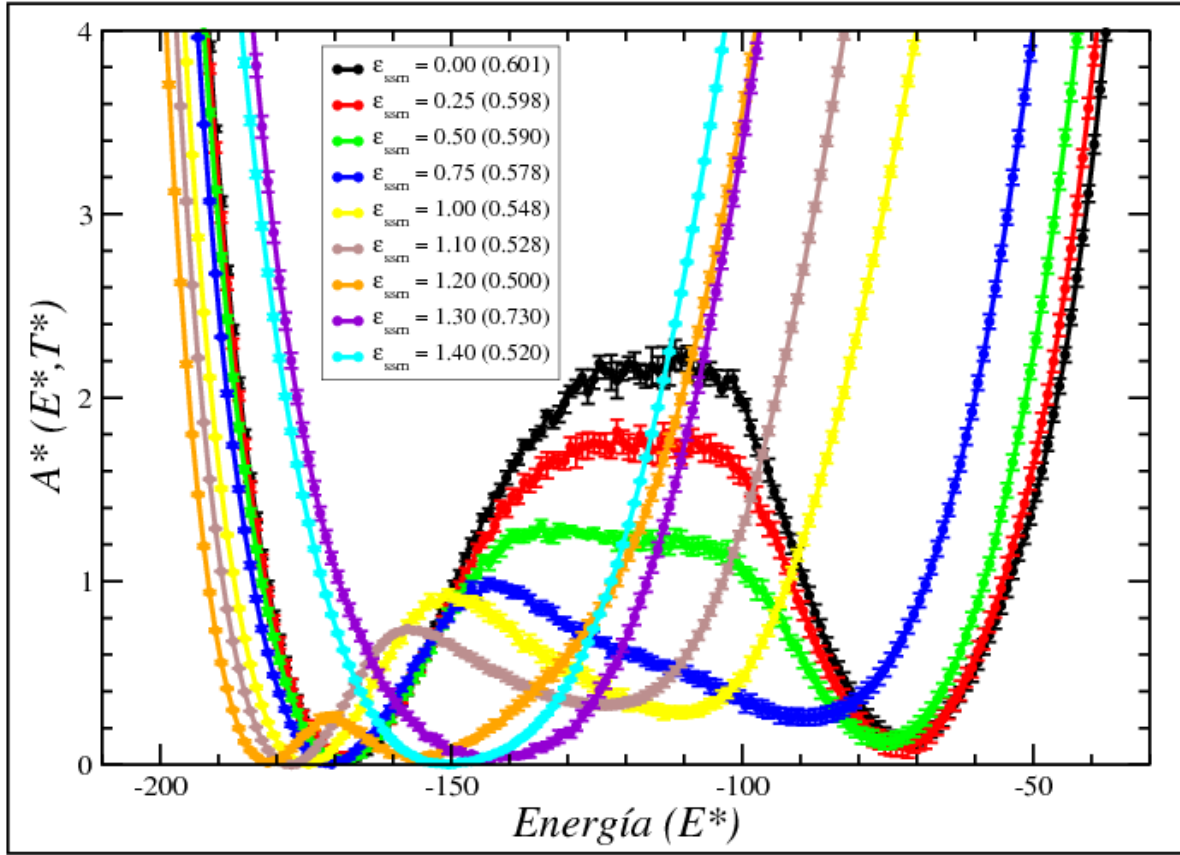


Figura 3.8: **Perfiles de energía de Helmholtz en función de la presión.** En la leyenda se reflejan los valores de ϵ_{ssm} para cada sistema. Entre paréntesis se muestra la temperatura de transición correspondiente a cada presión, a la que está realizado el perfil de energía de Helmholtz.

que observamos es un tránsito no cooperativo entre diferentes estados desnaturalizados. Dicho tránsito no muestra un perfil de doble mínimo en la representación de la energía de Helmholtz, pero sí muestra un pico ensanchado en la curva de capacidad calorífica. Por tanto, tendremos que estudiarlo para poder analizar qué fenómeno se está dando para altas presiones a la temperatura del máximo de la curva de capacidad calorífica.

- Por otro lado, observamos una clara evolución del estado desnaturalizado. A bajas presiones, hasta valores de $\epsilon_{ssm} = 0.75$, apenas existe una modificación en su energía, ya que todos los estados desnaturalizados presentan valores de energía superiores a -100 unidades reducidas (u.r.). Cuando $\epsilon_{ssm} \geq 0.75$, existe una clara evolución del estado desnaturalizado hacia valores menores de energía, siendo

la energía del estado desnaturalizado cuando $\epsilon_{ssm} = 1.20$ de aproximadamente -150 u.r. Esto denota una clara evolución del estado desnaturalizado que va más allá de las posibles fluctuaciones térmicas. Es como si el estado desnaturalizado evolucionara hacia un nuevo estado, al aumentar la presión, que en cualquier caso no coincide con el estado nativo.

- También es evidente que existe una clara evolución de la barrera de energía de Helmholtz que separa al estado nativo del estado desnaturalizado. Esta va disminuyendo considerablemente hasta hacerse casi nula cuando $\epsilon_{ssm} = 1.20$. Por tanto, observamos una nueva evidencia de que al aumentar la presión el tipo de tránsito entre el estado nativo y el desnaturalizado evoluciona desde un tránsito bastante cooperativo hasta un tránsito mucho menos cooperativo.

Para confirmar la existencia de un segundo tránsito en la simulación a altas presiones, tal y como sospechamos por el segundo pico en la curva de capacidad calorífica, estudiamos los perfiles de energía de Helmholtz a la temperatura correspondiente al segundo máximo en cada una de las simulaciones de mayor presión. En la Figura 3.9 mostramos los perfiles de energía para las simulaciones con $\epsilon_{ssm} = 1.00, 1.10, 1.20, 1.30$ y 1.40 en línea continua. Cada uno de estos perfiles está realizado a la temperatura de transición del segundo máximo de la curva de capacidad calorífica, que se señala en la leyenda. Podemos observar cómo estos perfiles muestran la existencia de un tránsito *downhill*, puesto que la forma de dicho perfil es característica de un tránsito sin barrera, que difiere al que habíamos observado a T_m (el primer pico en la curva de capacidad calorífica). Además, si atendemos a los valores de energía a los que aparece el perfil, corresponden a un valor intermedio entre el estado desnaturalizado de menor energía, propio de las simulaciones de alta presión (Figura 3.8) y el estado desnaturalizado de alta energía que observamos a presión atmosférica. En la Figura 3.9 se muestran, como comparación con línea discontinua, los perfiles de energía de Helmholtz cuando $\epsilon_{ssm} = 0.00$, en el intervalo de temperaturas donde aparece el segundo máximo en las simulaciones

de alta presión. De acuerdo con esta Figura queda claro que el estado desnaturalizado que se obtiene a altas presiones difiere notablemente en su valor de energía con respecto al estado desnaturalizado a presión atmosférica. Otro aspecto interesante de esta representación es el hecho que a medida que aumenta la presión, el estado que se estabiliza es de menor energía pese a encontrarse a una mayor temperatura. Este comportamiento es contrario a lo que habíamos observado habitualmente, donde un aumento de la temperatura implica un aumento de la energía del estado estabilizado. Esto nos hace pensar que a medida que aumenta la presión el sistema tiende a estabilizar un estado desnaturalizado de menor energía. Nos encontramos ante la existencia de **diferentes tipos de tránsito en función de los estados involucrados y de la presión a la que se encuentre el sistema.**

Por tanto, a través de las distintas curvas de capacidad calorífica y perfiles de energía de Helmholtz, que hemos obtenido con los diferentes valores de ϵ_{ssm} donde hemos simulado el sistema, hemos adquirido bastante información de cómo evoluciona el proceso de plegamiento para esta proteína en función de la presión a la que esté sometida. Hemos observado la evolución del tipo de tránsito, desde muy cooperativos cuando nos encontramos a presión atmosférica, hasta transitos continuos sin barrera cuando las presiones son mucho más altas.

$$\Downarrow \epsilon_{ssm} \propto \text{Tránsito cooperativo}$$

$$\Uparrow \epsilon_{ssm} \propto \text{Tránsito continuo}$$

También hemos podido observar cómo con nuestro modelo el efecto de la presión empieza a ser notable, al menos, cuando el valor de $\epsilon_{ssm} \geq 0.75$. Esto nos indica que el efecto más acusado de la presión aparece cuando el valor de ϵ_{ssm} se aproxima al valor de ϵ_{go} . Por otro lado, podemos hablar de presión muy alta cuando el valor de ϵ_{ssm} se hace superior a 1.2. Por tanto, nos encontramos con tres regímenes de presión bien diferenciados cuando estudiamos al plegamiento de proteínas con este modelo. Estos tres regímenes

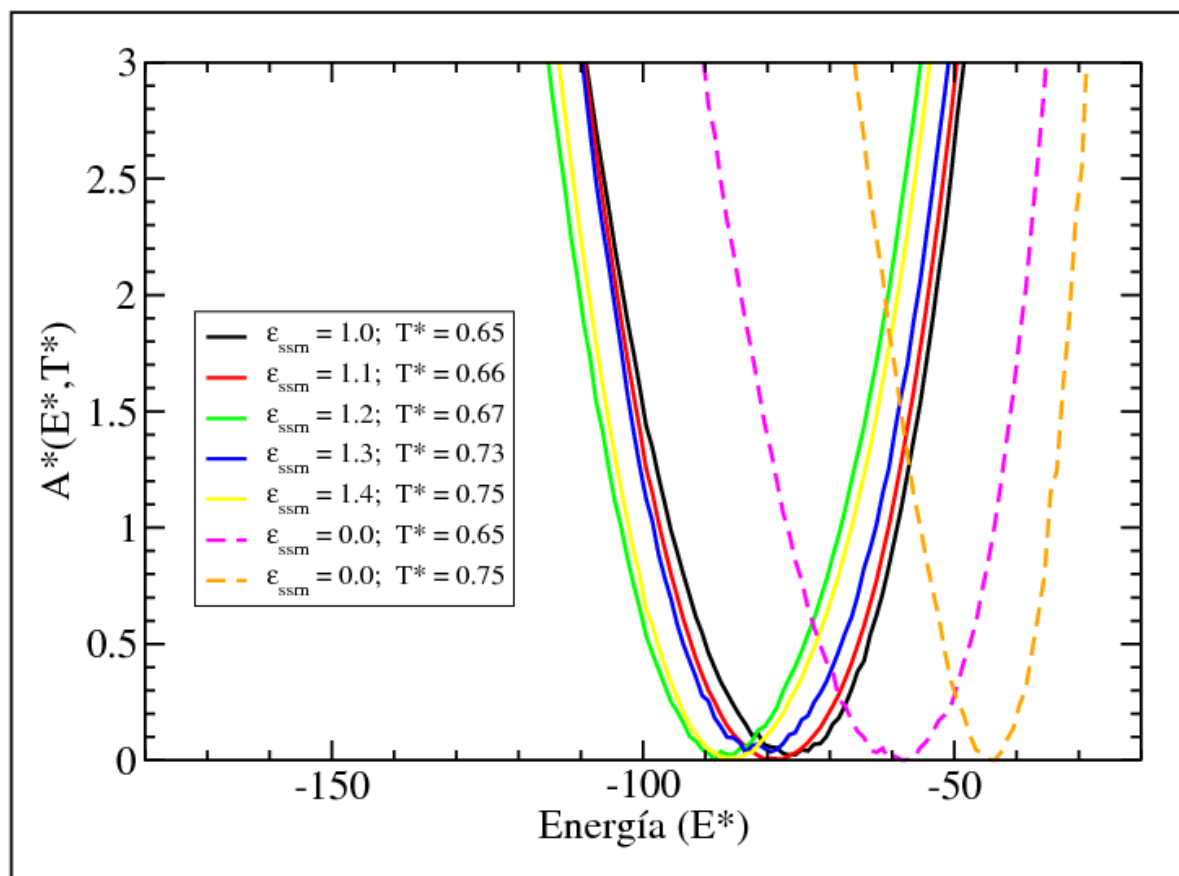


Figura 3.9: **Perfiles de energía de Helmholtz en función de la presión.** Se muestran los diferentes perfiles de energía de Helmholtz, a la temperatura a la que aparece el segundo máximo en la curva de capacidad calorífica (cuyo valor que se muestra en la leyenda), para los diferentes sistemas a presiones crecientes. Las dos curvas con línea discontinua muestran el perfil de energía de Helmholtz para la simulación a presión atmosférica, en el intervalo de temperaturas donde se representan el resto de perfiles de energía.

de presión podemos relacionarlos de una manera aproximada con los valores de presión que encontramos en la bibliografía para la desnaturalización de proteínas mediada por una presión hidrostática. A lo largo del Capítulo 5 desarrollaremos esta relación.

3.5.2. Evaluación de los diferentes estados involucrados en la transición

Para poder estudiar los diferentes estados encontrados a través de la simulación, y a la vez poder confirmar la aparición de un ‘nuevo’ estado desnaturalizado debido al efecto de la presión, nos plantemos el análisis estructural de las diferentes configu-

raciones obtenidas a lo largo de la simulación. Analizando los fotogramas que hemos obtenido a lo largo del proceso de plegamiento podemos estudiar la existencia o no de un estado distinto al estado nativo y al estado desnaturalizado ‘clásico’ debido a la temperatura, que es el que obtenemos a presión atmosférica. Para ello utilizamos dos de las propiedades diferentes que obtenemos de cada una de las configuraciones a través de las representaciones bidimensionales mencionadas anteriormente. En este punto decidimos descartar la energía como parámetro para clasificar cada uno de los estados. Esta decisión fue tomada por dos motivos principales: Por un lado, la existencia de diferentes temperaturas de transición en función de la presión a la que se encuentra sometido el sistema, ya que la energía accesible para cada sistema depende de la temperatura a la que se encuentre, y por tanto estados similares que obtengamos a diferentes temperaturas pueden ser diferentes en energía, por la distinta amplitud de sus fluctuaciones térmicas, lo cual complica notablemente la comparación entre sistemas que se encuentran a distintas presiones. En cualquier caso las fluctuaciones térmicas del sistema, que definimos en la Sección 3.4.1, serán un problema subyacente en muchas fases de nuestro trabajo, siempre que queramos comparar configuraciones que procedan de diferentes presiones o distinta temperatura. Por otro lado, como vimos con los perfiles de energía, el nuevo estado desnaturalizado que aparece por efecto de la presión puede llegar a tener energías muy similares al estado nativo, como se puede observar en la Figura 3.4 cuando $\epsilon_{ssm} = 1.20$. Esto también supone un problema de cara a clasificar distintos estados en función de su energía, ya que si estados diferentes tienen energías similares se hace excesivamente complejo diferenciar estos estados usando la energía como criterio. Por tanto, esta deja de ser la propiedad idónea para clasificar los distintos estados, y parece conveniente tomar otro tipo de propiedad para comparar las diferentes configuraciones. Decidimos usar dos propiedades estructurales que ya hemos descrito, el radio de giro y el RMSD respecto a la estructura nativa, para la comparación entre distintos estados, ya que tras realizar diferentes pruebas comprobamos que estas propiedades presentan

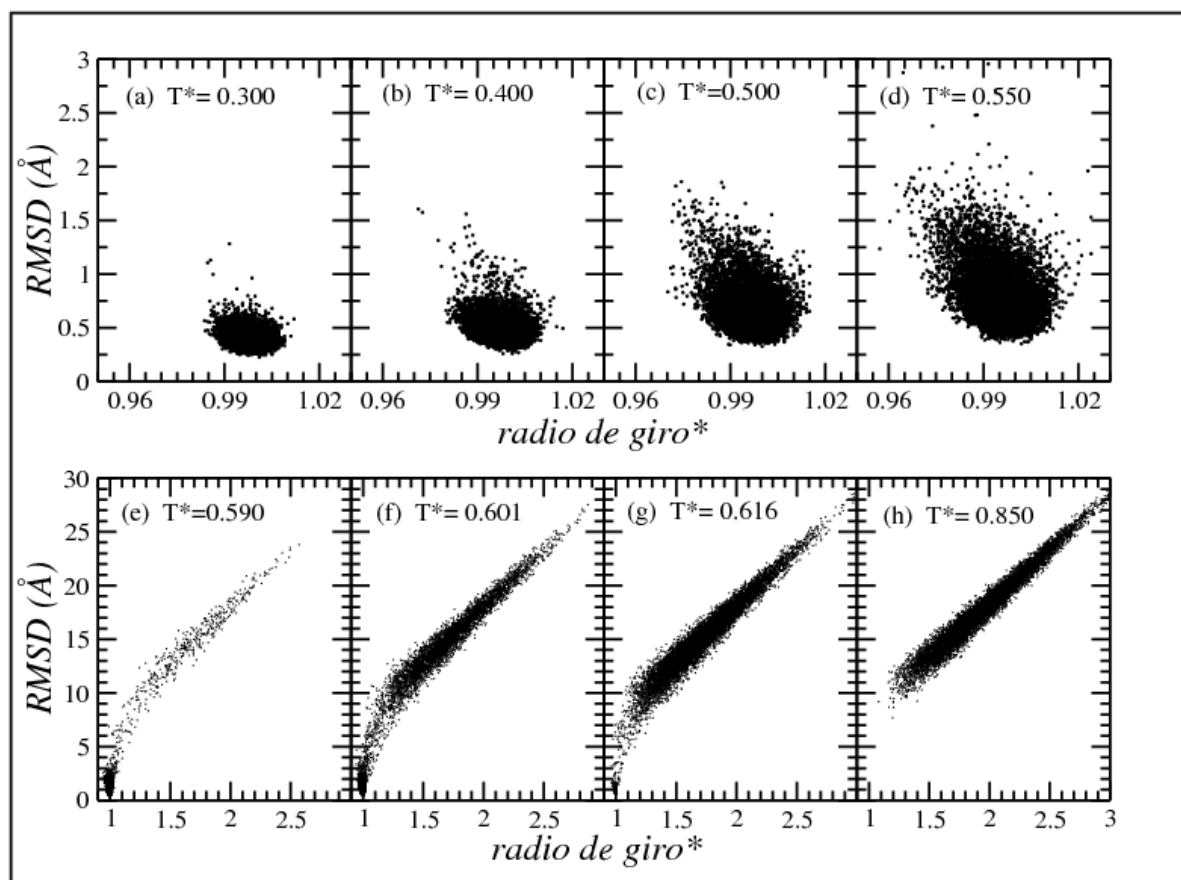


Figura 3.10: **Representación de las configuraciones, a presión atmosférica, a diferentes temperaturas.** Se muestran las diferentes configuraciones registradas a lo largo de la simulación en función de los diferentes valores de radio de giro reducido y RMSD. Se han seleccionado algunas de las temperaturas mas significativas para mostrar la evolución de los diferentes estados. La fila superior y la fila inferior se encuentran en diferentes escalas para que la comparación discutidas en el texto sea más sencilla. Todas las representaciones corresponden a un valor de $\epsilon_{ssm} = 0.00$.

cambios significativos a lo largo de las diferentes situaciones encontradas. Al mismo tiempo, no se ven tan influidos por la temperatura del sistema, sino que al tratarse de propiedades estructurales los cambios que se observan están relacionados con cambios en las estructuras, y estos son más fáciles de relacionar con la aparición de diferentes estados conformacionales. En la Figura 3.10 mostramos la representación bidimensional de las configuraciones obtenidas a presión atmosférica, $\epsilon_{ssm} = 0.0$, en función de su valor de radio de giro reducido y RMSD.

Hay que puntualizar que para facilitar la comparación de los resultados, hemos reducido el número de configuraciones que mostramos a cada temperatura hasta apro-

ximadamente 10000 configuraciones por representación, algo que haremos a lo largo de este Capítulo, puesto que un exceso de configuraciones dificulta el análisis visual de las imágenes, especialmente cuando tratamos de estimar la densidad de los diferentes estados. La selección de las configuraciones mostradas es completamente aleatoria, e igualmente distribuida a lo largo de la simulación. En esta Figura presentamos la evolución del muestreo a diferentes temperaturas. Hemos seleccionado algunas temperaturas características dentro del amplio rango estudiado: las cuatro primeras temperaturas, $T^* = 0.300, 0.400, 0.500$ y 0.550 , situadas en la fila superior de la imagen, paneles (a), (b), (c) y (d), respectivamente, corresponden a temperaturas por debajo de T_m para esta presión. Se observa cómo para estas temperaturas se estabiliza un único estado con valores bajos de radio de giro reducido y RMSD, inferiores a 1.02 y 2.50 Å, respectivamente. Se trata del estado nativo de la proteína. Si mostramos diferentes temperaturas donde sólo aparece dicho estado es para que se pueda observar el efecto de las fluctuaciones térmicas del sistema comentado anteriormente, ya que a medida que aumenta la temperatura el estado nativo aumenta ligeramente su valor promedio de radio de giro reducido y RMSD, lo que nos muestra la estrecha relación existente entre la temperatura del sistema y las posibles configuraciones que adopta la proteína dentro de un mismo estado. En la parte inferior de la imagen se muestran las configuraciones para temperaturas mayores. Nótese que existe un cambio apreciable de escala entre ambas filas de las gráficas. En esta segunda batería de imágenes se puede apreciar la aparición de un segundo estado, con valores mucho mayores de radio de giro reducido y RMSD, superiores a 1.25 y 8 Å, respectivamente. Cuando $T^* = 0.601$, panel (f), nos encontramos a la temperatura de transición y se observan claramente ambos estados en equilibrio. Aunque existen algunas conformaciones con valores de radio de giro reducido y RMSD intermedios, estos se observan en una población muy inferior a las configuraciones correspondientes a los dos estados principales, lo que confirma la existencia de una barrera energética considerable entre el estado nativo y el estado desnaturalizado, como había-

mos visto en la Figura 3.8. A $T^* = 0.616$ panel (g) se observa cómo el estado nativo se va despoblando de forma evidente para dar lugar exclusivamente al estado desnaturalizado. La última imagen, correspondiente a la temperatura más alta estudiada para esta presión, $T^* = 0.850$ panel (h), muestra cómo sólo se estabiliza el estado desnaturalizado, ya que no existen configuraciones con valores bajos de radio de giro reducido y RMSD. De esta forma queda de nuevo demostrada la existencia de un equilibrio de dos estados entre el estado nativo y el estado desnaturalizado a presión atmosférica para esta proteína, como también observamos en la representación bidimensional de la energía frente al número de contactos de la Figura 3.5 (b).

De un modo similar analizamos el resto de nuestras simulaciones. Si bien cada uno de los estudios a cada presión conlleva la representación de las diferentes conformaciones a cada temperatura, a continuación mostraremos algunas de las más relevantes observadas para algunas presiones a valores específicos de temperatura. En la Figura 3.11 mostramos las diferentes configuraciones obtenidas para diferentes presiones, siempre en función de sus valores de radio de giro reducido y RMSD. Dichas configuraciones las hemos obtenido a la correspondiente temperatura del máximo de la curva de capacidad calorífica para cada presión simulada. Analizando estas representaciones y comparándola con la Figura 3.10 (f), la que corresponde a la de más baja presión en esta serie, podemos observar cómo de nuevo existen tres tipos de comportamiento en función del valor de ϵ_{ssm} :

1. Para los valores más bajos de presión, hasta $\epsilon_{ssm} = 0.50$, paneles (a) y (b) de la Figura 3.11, observamos la presencia de dos estados diferenciados, del mismo modo que sucedía en la Figura anterior a presión atmosférica, si bien a medida que aumenta el valor de ϵ_{ssm} se observa cómo cada vez están más poblados los estados intermedios que aparecen entre el estado nativo y el estado desnaturalizado. Puesto que en la representación bidimensional no se aprecia perfectamente este efecto, al menos para la resolución que podemos ofrecer en esta imagen, la Figura 3.12 se

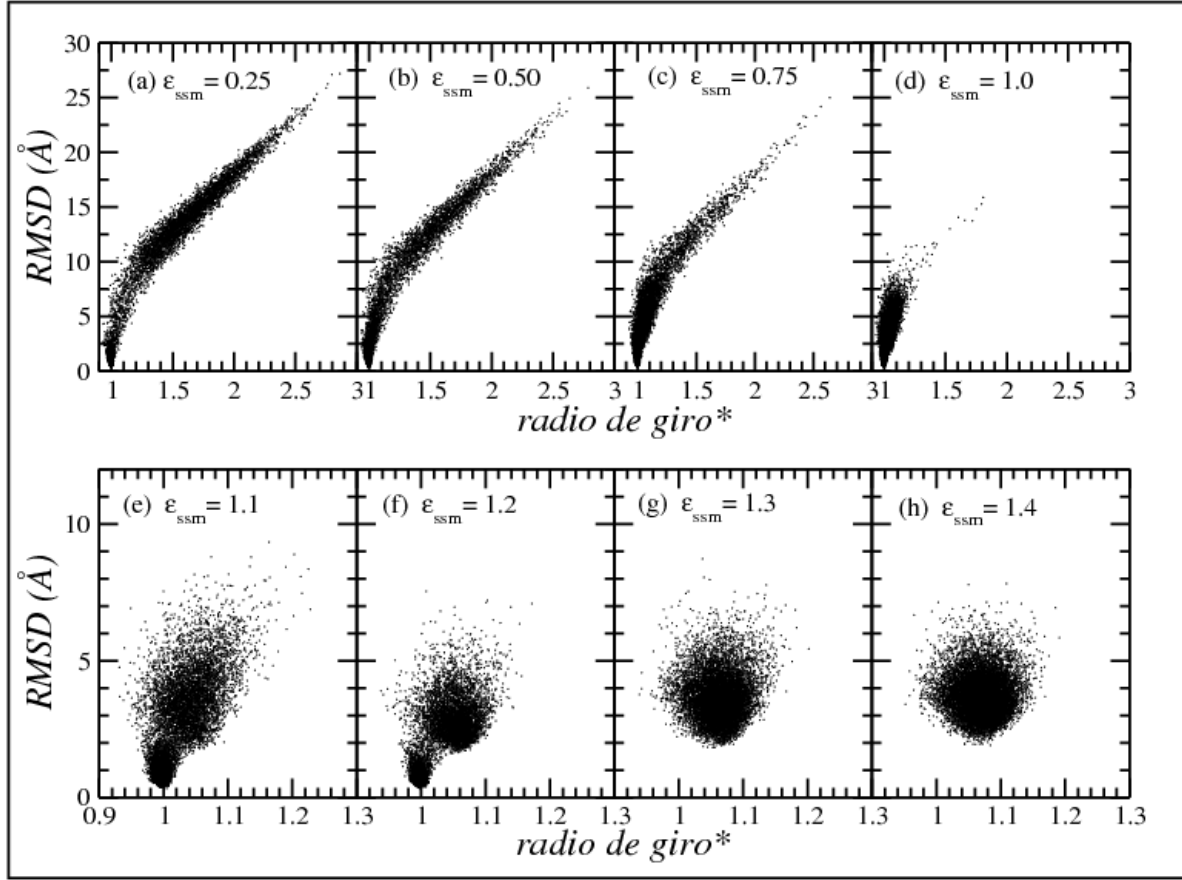


Figura 3.11: **Representación de las configuraciones, para valores crecientes de ϵ_{ssm} , a las diferentes temperaturas del máximo de la curva de capacidad calorífica.** Se muestran las diferentes configuraciones obtenidas a lo largo de la simulación en función de los diferentes valores de radio de giro reducido y RMSD.

muestran los diferentes histogramas resultantes para los valores de radio de giro reducido, en la fila superior, y RMSD en la fila inferior, que se han obtenido para todas las configuraciones registradas. En los histogramas que mostramos, para los valores de $\epsilon_{ssm} = 0.00, 0.25, 0.50$ y 0.75 , siempre a sus temperaturas de transición, queda perfectamente patente cómo un aumento de la presión implica el aumento de población de las configuraciones con valores intermedios de radio de giro reducido y RMSD, pasándose de dos máximos bien diferenciados cuando $\epsilon_{ssm} = 0.00$, Figura 3.12 (a) y (e), a prácticamente un continuo en el caso en que ϵ_{ssm} vale 0.50 , Figura 3.12 (c) y (g). En esta imagen también queda muy bien reflejado cómo cuando el valor de $\epsilon_{ssm} = 0.75$, el comportamiento es claramente diferenciado al que encontramos a presiones más bajas.

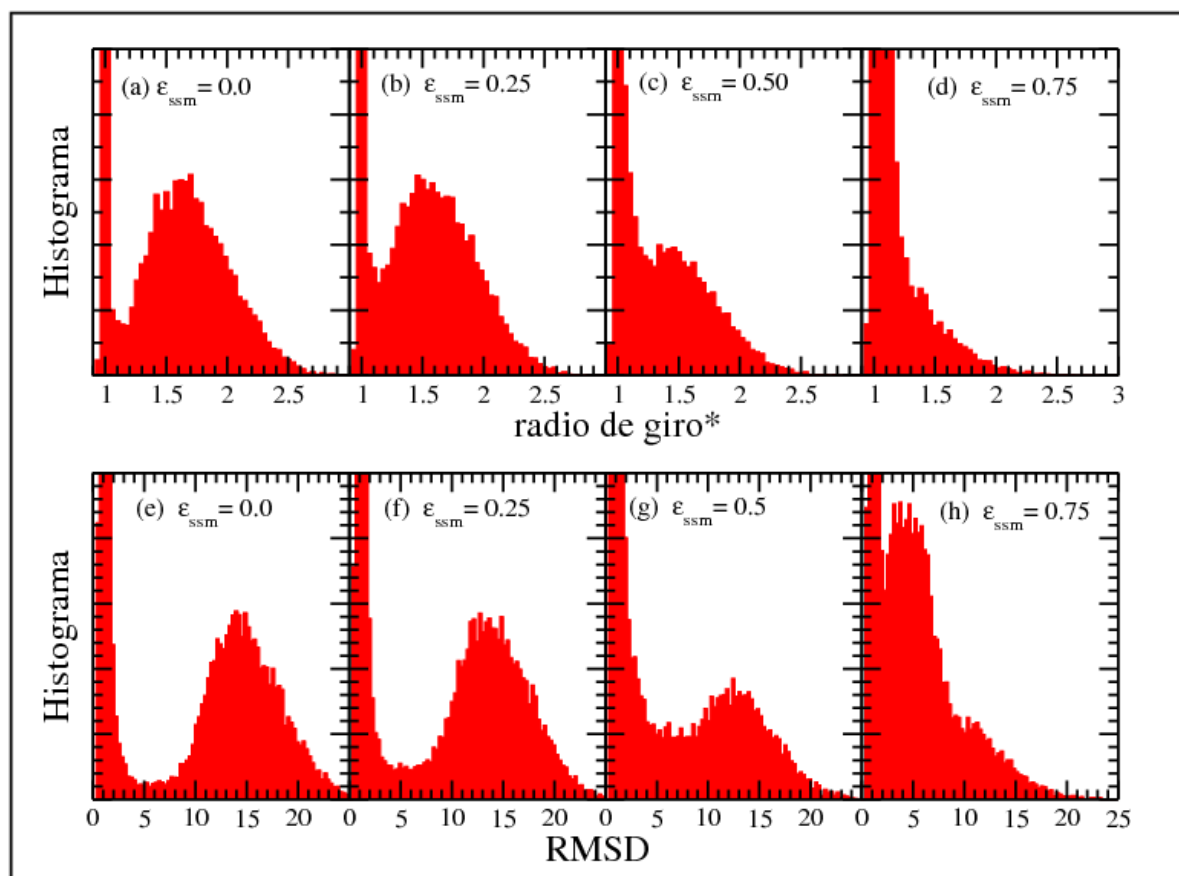


Figura 3.12: **Histogramas de los valores de radio de giro reducido y RMSD, para las simulaciones de baja presión, a la temperatura de transición.** Se muestran los histogramas de las dos propiedades utilizadas para la clasificación de configuraciones, radio de giro reducido y RMSD, para las simulaciones de bajo valor de ϵ_{ssm} a su temperatura de transición.

2. A presiones moderadas, cuando ϵ_{ssm} vale entre 0.75 y 1.20, Figuras 3.11 (c-f), se observa cómo el estado desnaturalizado comienza a evolucionar progresivamente hacia menores valores de radio de giro reducido y RMSD. Este fenómeno, que sucedía también a presiones bajas de forma muy leve, se ve fuertemente acentuado a partir de estos valores de ϵ_{ssm} . Esto refleja un cambio estructural del estado desnaturalizado a tenor del desplazamiento de valores de radio de giro reducido y RMSD, un comportamiento que es especialmente notable cuando el valor de ϵ_{ssm} es igual a 0.75 y 1.00. Por último, cuando $\epsilon_{ssm} = 1.10$ y 1.20, Figuras 3.11 (e) y (f), el nuevo estado desnaturalizado queda bastante más estabilizado. Lo que se observa en estas gráficas son dos estados muy cercanos en valores de radio de giro

reducido y RMSD, uno correspondiente al estado nativo, siempre presente hasta estos valores de presión, y el nuevo estado desnaturalizado con valores de radio de giro reducido y RMSD muy inferiores al estado desnaturalizado que observábamos a presión atmosférica.

3. Por último, a altas presiones, podemos ver cómo para valores de $\epsilon_{ssm} = 1.30$ y 1.40 , Figuras 3.11 (g) y (h), existe un único estado. Ya no existe el estado nativo tal y como veníamos observando en la serie, sino que únicamente se estabiliza el estado desnaturalizado con bajos valores de radio de giro reducido y RMSD. Por tanto, a altas presiones se confirma la no existencia del estado nativo, al menos a la temperatura del máximo de la curva de capacidad calorífica. Recordemos que, a altas presiones, este es un máximo muy poco definido (Figura 3.6), que no parece implicar una transición entre diferentes estados, sino un único estado desnaturalizado. Esto vuelve a corroborar que el análisis de las curvas de capacidad calorífica y los perfiles de energía no son absolutos, sino que deben ir acompañados de un minucioso análisis de las propiedades estructurales.

Por otro lado, nos parece interesante mostrar la evolución completa de una de las simulaciones de presión moderada a lo largo de diferentes temperaturas. Se trata por tanto de observar cómo para un valor seleccionado de ϵ_{ssm} van apareciendo los diferentes estados en función de la temperatura a la que nos encontramos. Seleccionamos un valor de ϵ_{ssm} igual a 1.20 por resultar especialmente revelador para el proceso de desnaturalización debido a la presión. En la Figura 3.13 se muestran las características de las diferentes configuraciones obtenidas para este valor de ϵ_{ssm} . A medida que aumenta la temperatura se estabilizan hasta tres estados distintos. A bajas temperaturas, el único estado estable es el **estado nativo**, al que nos referiremos también como **estado N**. A medida que aumenta la temperatura se va apreciando la aparición de un segundo estado. Este primer estado desnaturalizado, con bajo valor de radio de giro reducido y RMSD, se trata además de un estado desnaturalizado de baja energía, ya que corresponde al estado

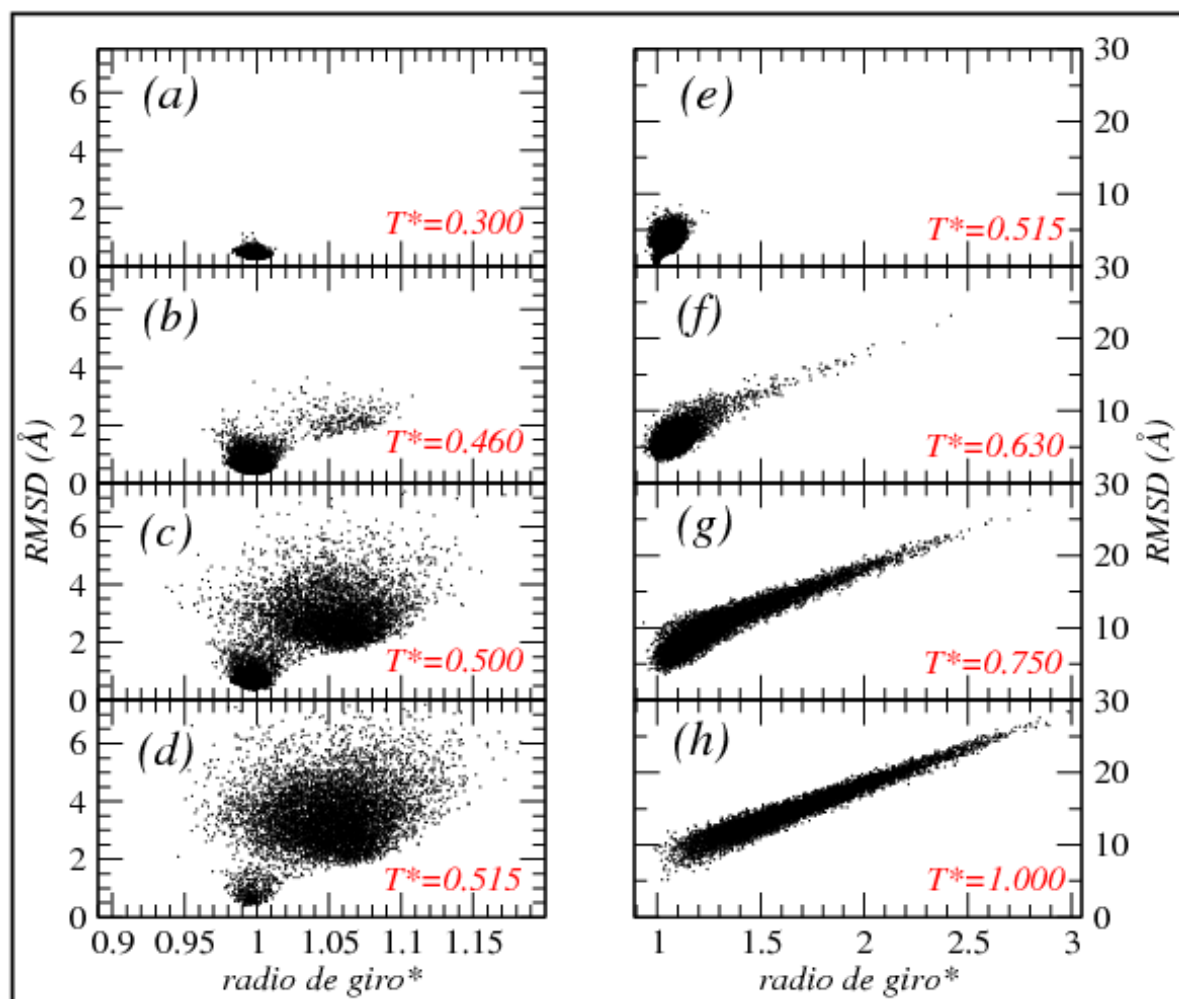


Figura 3.13: **Evolución de las distintas configuraciones a presión moderada, a medida que aumenta la temperatura.** Se muestran las distintas configuraciones obtenidas en función del valor de radio de giro reducido y RMSD cuando $\epsilon_{ssm} = 1.20$. Las temperaturas seleccionadas en esta Figura son características de la evolución observada a esta presión. Nótese que las imágenes d) y e) son exactamente la misma pero con diferente escala.

desnaturalizado que se observaba en el perfil de energía de Helmholtz para $\epsilon_{ssm} = 1.20$ de la Figura 3.8. Este estado, que corresponde a un estado desnaturalizado debido al efecto de la presión, es un estado bastante compacto a tenor de sus bajos valores de radio de giro reducido y RMSD, que a partir de ahora denominaremos **estado hinchado** o **estado H**. Además, a la temperatura del máximo de la curva de capacidad calorífica para esta presión, $T^* = 0.500$, cuyas configuraciones aparecen representadas en la Figura 3.13 (c), aunque coexisten ambos estados estos son muy similares, y apenas se aprecia la ausencia de estados intermedios entre ambos máximos, lo que corrobora que la barrera energética

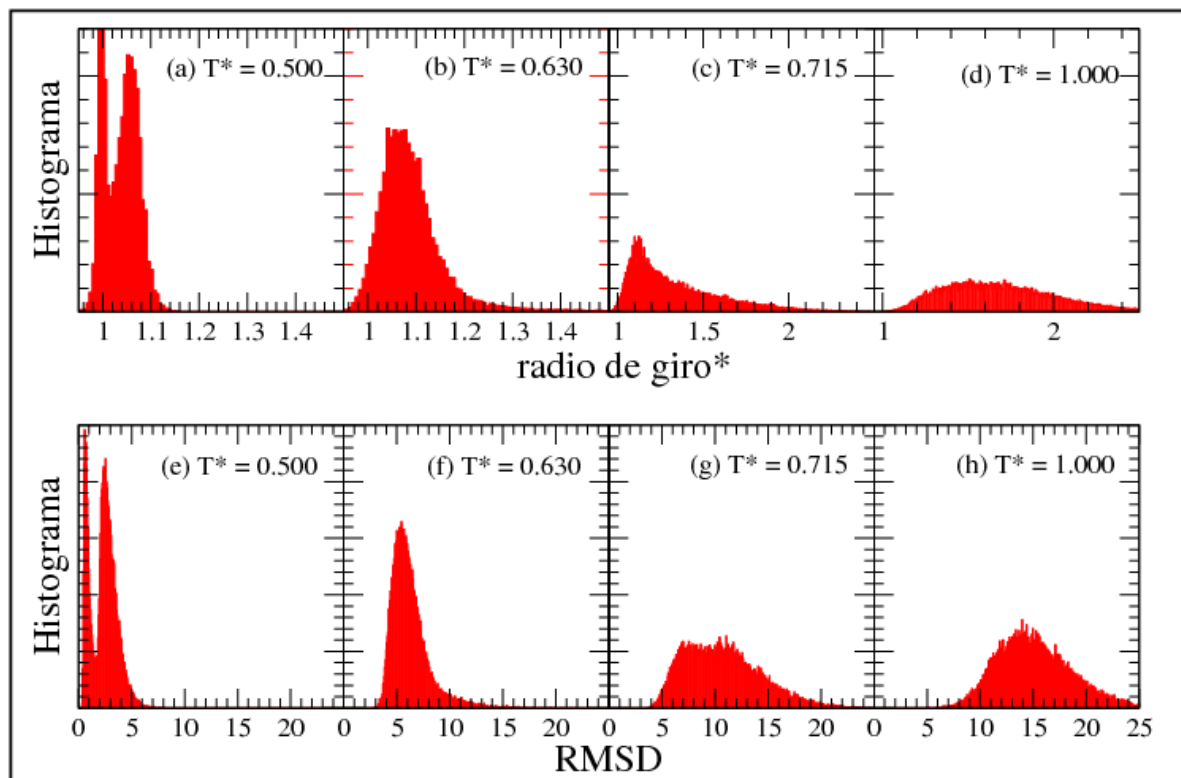


Figura 3.14: **Histogramas del radio de giro reducido y RMSD para $\epsilon_{ssm} = 1.20$, a diferentes temperaturas.** Se muestran los diferentes histogramas obtenidos para las simulaciones de presión moderada cuando $\epsilon_{ssm} = 1.20$. Se han seleccionado algunas de las temperaturas características de los tránsitos encontrados a esta presión. Las temperaturas que aquí se muestran son algunas de las que aparecían en la Figura 3.13.

que se observaba en su perfil de energía de Helmholtz fuese muy baja. Esto nos permite especular con un tránsito de dos estados poco cooperativo, entre el estado N y el estado H, o bien con un tránsito que mostraría carácter de tipo *downhill*. En cualquier caso, la barrera de energía entre ambos estados es prácticamente inexistente, característica que otros grupos de investigación han atribuido al tránsito entre el estado nativo y el estado hinchado, por tratarse de un tránsito que se produce en un intervalo de tiempo muy breve¹⁶⁵. Hay que mencionar que las Figuras 3.13 (d) y (e) son exactamente la misma con la diferencia de un cambio de escala para una mejor comparación con el resto de imágenes.

A partir de la temperatura del máximo de la curva, observamos no sólo la progresiva desaparición del estado nativo, sino que además existe una evolución paulatina del estado hinchado hacia un estado desnaturalizado con valores altos de radio de giro

reducido y RMSD. Este estado desnaturalizado, que ya no es compacto, es similar en valores de radio de giro, RMSD y energía al estado desnaturalizado que encontramos a presión atmosférica, el **estado desnaturalizado** ‘clásico’ debido a la temperatura, que denominaremos **estado D** para diferenciarlo del estado desnaturalizado debido a la presión. Se observa claramente cómo el tránsito entre los dos estados desnaturalizados, H y D, es claramente no cooperativo, puesto que la evolución desde un estado hacia el otro se da de forma progresiva y sin ningún tipo de barrera poblacional. Esta situación concuerda con los perfiles de energía de Helmholtz que habíamos mostrado anteriormente, a la altura del segundo máximo de C_v , donde no se observaba barrera de energía entre diferentes estados (Figura 3.9). Para confirmar la existencia o no de configuraciones intermedias entre los diferentes estados clasificados, en la Figura 3.14 mostramos los histogramas obtenidos para los valores de radio de giro reducido y RMSD para algunas de las temperaturas características de esta transición. En las gráficas 3.14 (a) y (e), con $T^* = 0.500$, se observa claramente la coexistencia de los dos estados, N y H, separados por una pequeña barrera poblacional. Cuando la temperatura es superior a T_m , Figuras 3.14 (b-d, f-h), observamos la presencia de un único estado que evoluciona con la temperatura, pero en ningún caso se aprecia la aparición de dos máximos, lo que nos confirma que nos encontramos ante una transición continua entre el estado hinchado y el estado desnaturalizado.

Por último, y para obtener la información completa de lo que sucede a todas las presiones estudiadas, mostramos las representaciones bidimensionales de las configuraciones obtenidas a altas presiones. Así en la Figura 3.15 observamos los resultados para nuestra simulación de mayor presión, $\epsilon_{ssm} = 1.40$, a diferentes temperaturas. Si bien en dicha Figura se puede observar cómo en ningún momento se estabiliza más de un estado diferente a la misma temperatura, sí que existen algunas consideraciones importantes. En primer lugar hay que notar que incluso a la temperatura más baja, $T^* = 0.300$, Figura 3.15 (a), solo se estabiliza el estado hinchado, sin que exista estado nativo en

ningún momento. Nos encontramos en un régimen de alta presión que no permite a la proteína estabilizar al estado nativo ni siquiera a las temperaturas más bajas estudiadas. A medida que aumenta la presión, el estado hinchado va evolucionando hacia el estado desnaturalizado aumentando su valor de radio de giro reducido y RMSD, como se puede observar en la Figura 3.15 (b) y (c). Esto se debe a las fluctuaciones térmicas del sistema, al igual que le sucede al estado nativo a presiones bajas y moderadas. A partir de $T^* = 0.500$, que corresponde al primer máximo en la curva de capacidad calorífica para esta presión, el estado hinchado se va desestabilizando para iniciar un tránsito continuo que nos lleva al estado D, Figuras 3.15 (d-h). Es el mismo tránsito continuo que observábamos para el régimen de presión moderada entre el estado H y el estado D, un tránsito sin barreras energéticas que permite poblar a lo largo de la transición los estados intermedios entre ambos estados.

De esta manera, hemos mostrado un análisis completo de todos los tránsitos y todos los estados que hemos ido encontrando a lo largo de las diferentes simulaciones que hemos realizado para esta proteína a distintas presiones.

3.6. Estudio de poblaciones de los diferentes estados

Una vez identificados tres estados diferentes en función de la temperatura y el valor del parámetro ϵ_{ssm} que toma nuestro sistema, decidimos realizar un estudio cuantitativo de la evolución de las poblaciones que encontramos para cada uno de estos estados con respecto a la temperatura. De esta manera podremos identificar la aparición y desaparición de los diferentes estados en función de la temperatura y la presión a la que se encuentra el sistema, y así poder entender dónde aparecen las ‘fronteras’ que caracterizan las diferentes transiciones involucradas.

Si bien el estudio de las poblaciones puede parecer en principio bastante sencillo de llevar a cabo, en este caso vino acompañado de una dificultad añadida, **la clasificación de todas las configuraciones del muestreo en uno de los tres estados**

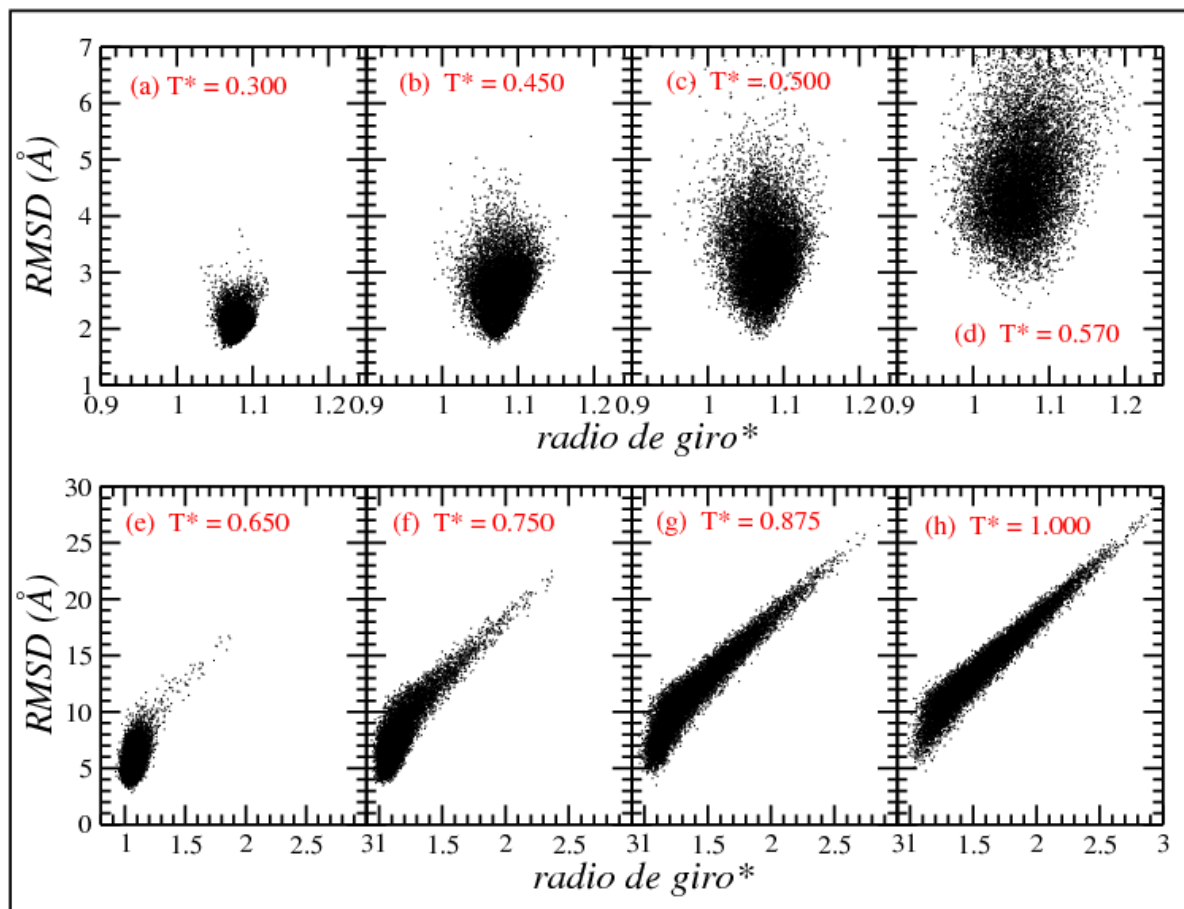


Figura 3.15: **Representación de las configuraciones cuando el sistema se encuentra a altas presiones.** Se muestran los diferentes configuraciones obtenidas para las simulaciones de alta presión cuando $\epsilon_{ssm} = 1.40$. Se han seleccionado algunas de las temperaturas características del tránsito que se observa entre el estado hinchado y el estado desnaturalizado.

descritos. A lo largo de la Sección anterior hemos observado la aparición de tres estados diferentes, estado nativo, hinchado y desnaturalizado. Al mismo tiempo hemos descrito las propiedades y características de cada uno de estos tres estados. Sin embargo, siempre existen estados intermedios, incluso en el paso desde el estado nativo al estado desnaturalizado, que se alejan en mayor o menor medida de cualquiera de los tres estados de referencia. Esta situación será especialmente grave en el caso de las transiciones entre el estado nativo y el hinchado y entre el estado hinchado y el estado desnaturalizado, ya que como hemos podido ver en estos dos casos las configuraciones intermedias son mucho más numerosas, al tratarse de tránsitos continuos.

Además, puesto que queremos realizar este estudio para cada una de las presiones

estudiadas, necesitamos encontrar parámetros de clasificación que sean igualmente válidos a todas las presiones a las que trabajamos. Para ello realizamos un estudio riguroso de las representaciones bidimensionales de todas las configuraciones en función de los valores de radio de giro reducido y RMSD, para así poder clasificar cada configuración en función del valor que toman estas propiedades. En la Figura 3.16 (a) mostramos la representación conjunta de todas las conformaciones, a la temperatura de transición, para la simulación donde $\epsilon_{ssm} = 0.00$ en rojo y para $\epsilon_{ssm} = 1.20$ en negro. La Figura 3.16 (b) muestra una ampliación de la zona de menores valores de radio de giro reducido y RMSD, donde se solapa el estado nativo de ambas simulaciones. Las Figuras 3.16 (c-f) muestran los diferentes histogramas que obtenemos al estudiar cada una de las propiedades estructurales por separado, en la parte superior se muestran los referentes al radio de giro reducido, mientras que en la parte inferior se muestran los histogramas de RMSD. En estos histogramas, las curvas en rojo corresponden de nuevo a la simulación con $\epsilon_{ssm} = 0.00$ y en negro a $\epsilon_{ssm} = 1.20$.

En la representación bidimensional se muestran conjuntamente los tres estados característicos, puesto que en rojo se observa el equilibrio entre el estado nativo y el desnaturalizado a baja presión, y en negro el tránsito entre el estado nativo y el estado hinchado a presión moderada. Aprovechando esta representación conjunta, señalamos con una línea discontinua azul los intervalos de radio de giro reducido y RMSD que hemos seleccionado como representativos del estado nativo, y con una línea discontinua verde los intervalos seleccionados para el estado desnaturalizado. Debe quedar claro que la selección de estos intervalos no se ha dictaminado únicamente a partir de estas dos representaciones, sino que la selección de los intervalos se realiza tras observar todas las simulaciones y los diferentes estados presentes, ya que como comentamos anteriormente las fluctuaciones térmicas del sistema hacen que en función de la temperatura a la que se encuentre el sistema estados iguales tengan valores algo diferentes de radio de giro reducido y RMSD. Los intervalos característicos de cada estado son los siguientes:

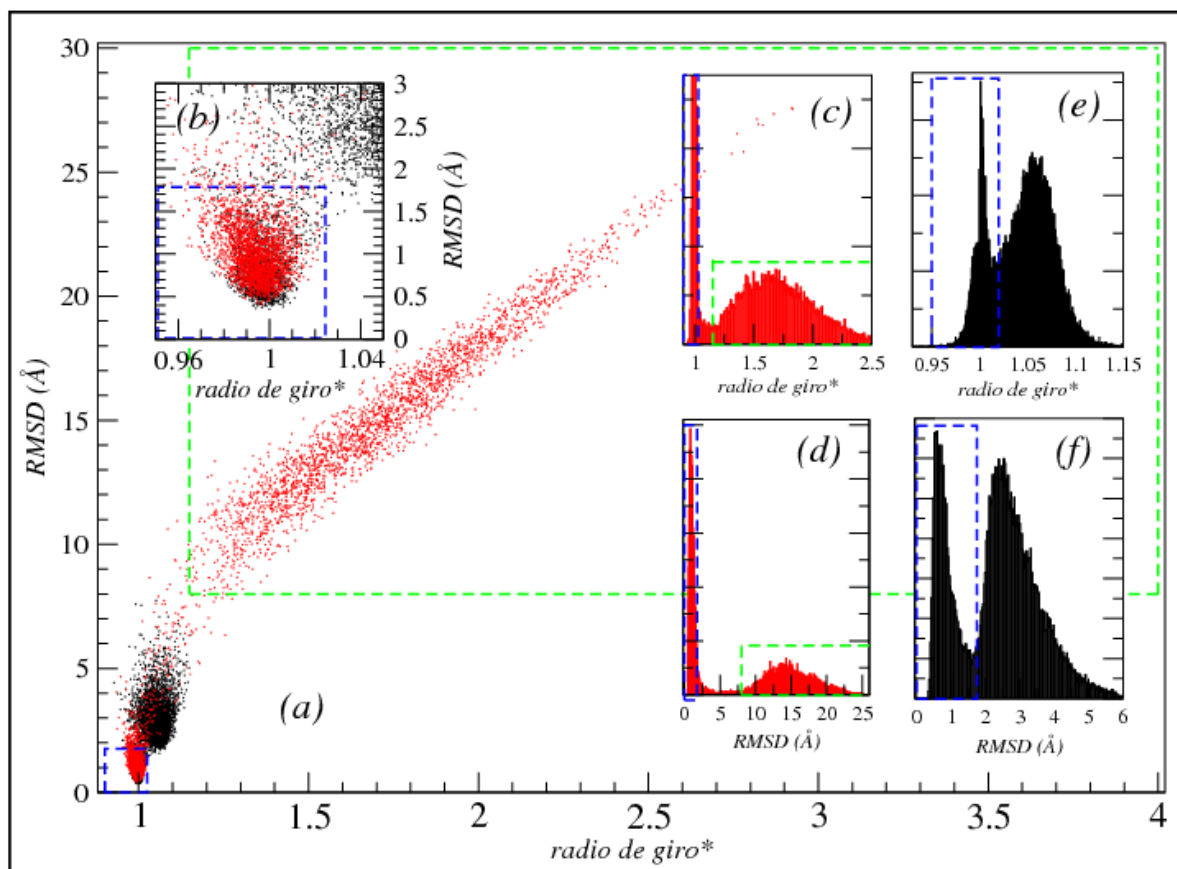


Figura 3.16: **Selección de las propiedades que identifican a cada estado.** a) Representación de todas las configuraciones presentes a la temperatura de transición cuando $\epsilon_{ssm} = 0.00$ (color rojo) y cuando $\epsilon_{ssm} = 1.20$ (color negro). b) Ampliación de a) en la zona de menores valores de radio de giro reducido y RMSD. c) Histograma de radio de giro reducido obtenido cuando $\epsilon_{ssm} = 0.00$. d) Histograma de RMSD obtenido cuando $\epsilon_{ssm} = 0.00$. e) Histograma de radio de giro reducido obtenido cuando $\epsilon_{ssm} = 1.20$. f) Histograma del RMSD obtenido cuando $\epsilon_{ssm} = 1.20$. Los rectángulos con línea discontinua seleccionan a lo largo de todas las gráficas los rangos seleccionados para los diferentes estados, en azul el estado nativo, en verde el estado desnaturalizado, en la zona intermedia entre ambos queda el estado hinchado.

- Estado nativo (recuadro de color azul), $R_g^* < 1.025$ y $\text{RMSD} < 1.8 \text{ \AA}$.
- Estado desnaturalizado (recuadro de color verde), $R_g^* > 1.150$ y $\text{RMSD} > 8.0 \text{ \AA}$.
- Estado hinchado (zona intermedia), $1.025 < R_g^* < 1.150$ y $1.8 \text{ \AA} < \text{RMSD} < 8.0 \text{ \AA}$.

La clasificación se ha hecho a partir del estado nativo y del estado desnaturalizado, puesto que ambos son los estados bien conocidos en el proceso de plegamiento de proteínas, y se encuentran presentes siempre a la temperatura de transición hasta presiones moderadas. Por tanto, se han seleccionado los valores que identifican a estos dos estados,

y se ha planteado el tercer estado, el estado hinchado, como el estado intermedio entre ambos.

Una vez elegido un criterio para la clasificación de cada uno de los estados, se realizó un análisis de poblaciones en función de la temperatura. En la Figura 3.17 se muestra el porcentaje de cada una de las poblaciones obtenido en función de la temperatura para cada una de las distintas presiones simuladas. De este modo podemos observar la evolución de estos tres estados en función del valor de ϵ_{ssm} , y por tanto en función de la presión. Analizando la Figura de menor a mayor valor de ϵ_{ssm} observamos diferentes aspectos interesantes de señalar. A bajos valores de presión, $\epsilon_{ssm} \leq 0.50$, notamos cómo las poblaciones demuestran la presencia significativa de solo dos especies, el estado nativo y el estado desnaturalizado de color negro. Si analizamos estos diagramas de poblaciones de menor a mayor temperatura observamos cómo a bajas temperaturas el estado nativo es el único poblado. A una temperatura determinada, $T^* \approx 0.6$, existe un cambio brusco y se estabiliza el estado desnaturalizado. Este es el comportamiento habitual para la transición de una proteína de dos estados. Además, en estos primeros diagramas de poblaciones, con $\epsilon_{ssm} \leq 0.50$, se puede observar cómo las pendientes de las dos curvas (estado nativo y desnaturalizado) son muy pronunciadas a la temperatura de transición, algo característico de los equilibrios de tipo todo o nada, donde en un rango muy estrecho de temperaturas encontramos un cambio muy brusco de configuraciones.

Cuando ϵ_{ssm} alcanza el valor 0.75, observamos cómo al menos en un pequeño intervalo de temperaturas el estado hinchado resulta ser el más poblado. En este caso se observa cómo ya no son sólo el estado nativo y el desnaturalizado los que participan en el proceso de desnaturalización, sino que a una temperatura algo inferior a 0.6 el estado hinchado resulta ser no sólo el más poblado sino que casi el 60 % de las configuraciones corresponden a este estado. A medida que el valor de ϵ_{ssm} sigue creciendo, podemos observar cómo la estabilización del estado hinchado va en aumento, no sólo porque sea el más poblado durante un intervalo mayor de temperaturas, sino porque el estado hinchado se convierte en el único estado razonablemente poblado dentro de un rango amplio

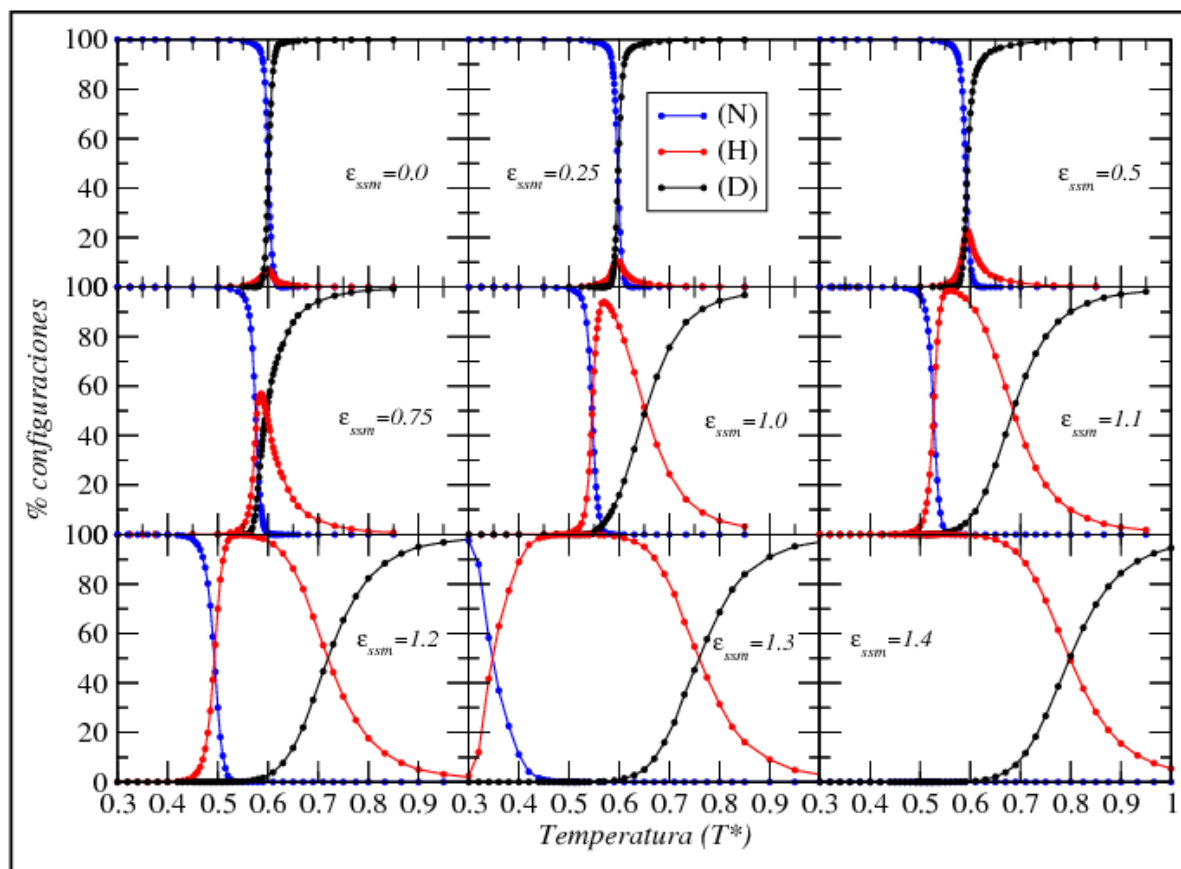


Figura 3.17: **Diagramas de poblaciones.** Mostramos la evolución de la población de los tres estados de referencia, en azul el estado nativo (N), en rojo el estado hinchado (H) y en negro el estado desnaturalizado (D), para los diferentes valores de ϵ_{ssm} en función de la temperatura.

de temperaturas. La estabilidad creciente del estado hinchado provoca que cuando $\epsilon_{ssm} \geq 0.75$ encontremos dos transiciones a lo largo del proceso de plegamiento, una primera transición entre el estado nativo y el estado hinchado, y una segunda transición entre el estado hinchado y el estado desnaturalizado. Además, gracias a la pendiente de las curvas de población, podemos notar cómo el tránsito entre el estado nativo y el hinchado es un tránsito más cooperativo que el tránsito entre el estado hinchado y el desnaturalizado, donde las pendientes son más pequeñas. Este comportamiento con la presencia de tres estados en función de la temperatura a la que nos encontremos se repite hasta que $\epsilon_{ssm} = 1.20$. En todo este intervalo de presiones se puede observar cómo el estado hinchado pasa a ser un estado muy estable durante un intervalo de temperaturas cada vez mayor.

Cuando $\epsilon_{ssm} = 1.30$, observamos un claro cambio de tendencia. Si bien a la temperatura más baja simulada el estado nativo sigue siendo el más estable, deja de serlo a muy bajas temperaturas, dando lugar al estado hinchado, que posteriormente evoluciona hasta el estado desnaturalizado. De alguna forma el estado nativo mantiene su estabilidad en un intervalo mucho más pequeño de temperaturas, lo que implica que los valores altos de ϵ_{ssm} desestabilizan claramente al estado nativo. Situación que se observa a altas presiones queda muy bien reflejado para $\epsilon_{ssm} = 1.40$, donde el estado nativo desaparece totalmente, en este caso ni siquiera a las menores temperaturas simuladas aparece poblado dicho estado, sino que desde las temperaturas más bajas es el estado hinchado el único presente, y posteriormente evoluciona hacia el estado desnaturalizado. Por tanto, para esta presión el estado nativo es totalmente inexistente para valores razonables de temperatura, y únicamente se observa población de los estados hinchado y desnaturalizado.

De esta manera hemos abarcado un estudio completo de la dependencia de los tres estados de referencia con respecto a la temperatura y al parámetro ϵ_{ssm} . Si recurrimos a nuestra clasificación de la presión en tres diferentes regímenes, podemos resumir cómo en función del régimen en el que nos encontremos nos encontramos ante una situación diferente:

- A bajas presiones observamos únicamente la presencia de los estados nativo y desnaturalizado en equilibrio, con una ausencia casi total del estado hinchado. Esta situación corresponde al comportamiento típico de una proteína de dos estados a presión atmosférica.
- A presiones moderadas observamos la presencia de los tres estados. A bajas temperaturas el más estable es el estado nativo, que a través de lo que parece un tránsito de dos estados, con una cooperatividad que decrece al aumentar la presión, da lugar al estado hinchado, que se va haciendo más estable cuanto mayor sea la presión. El estado hinchado a su vez da lugar a mayor temperatura a un tránsito

que conduce al estado desnaturalizado. En este régimen de presiones es posible por tanto aislar esencialmente un único estado a una determinada temperatura. Esta situación es factible sólo dentro de este régimen de presiones.

- A altas presiones se observa cómo el estado nativo no es estable, sino que ya desde las temperaturas más bajas se estabiliza el estado hinchado, que a través de una transición gradual da lugar al estado desnaturalizado.

3.6.1. Diagrama de fases ϵ_{ssm} - T con nuestro modelo

Otro aspecto importante que obtenemos de los estudios de poblaciones es la existencia de dos temperaturas de transición en las simulaciones donde el valor de $\epsilon_{ssm} \geq 0.75$. En la Tabla 3.1 mostramos las temperaturas de transición obtenidas para cada valor de ϵ_{ssm} . La temperatura característica de cada tránsito la obtenemos de la temperatura donde se cortan las curvas correspondientes a cada estado en la Figura 3.17. La Tabla muestra la temperatura a la que se ha dado el tránsito y el tipo de tránsito que hemos observado en función del estudio de poblaciones. Así, el tránsito (N-D) a la transición entre el estado nativo y el desnaturalizado, el tránsito (N-H) al tránsito entre el estado nativo y el hinchado, y por último el tránsito (H-D) esta referido al tránsito entre el estado hinchado y el desnaturalizado. Cuando en la Tabla no aparece valor alguno es que no hemos observado dicho tránsito en el diagrama de poblaciones para ese valor de ϵ_{ssm} .

Lo primero que queda patente al observar estas temperaturas de transición es que no en todos los casos coinciden los resultados del análisis de poblaciones con la temperatura del máximo de la curva de capacidad calorífica que habíamos obtenido en la Sección 3.5.1 (Figuras 3.6 y 3.7). Precisamente para que pueda hacerse una mejor comparación, en las dos últimas columnas de la Tabla 3.1 se recogen también las temperaturas características de los máximos que obtuvimos en las curvas de capacidad calorífica para cada valor de ϵ_{ssm} . Comparando las diferentes columnas, observamos que el primer máximo

ϵ_{ssm}	Tráns. N-D	Tráns. N-H	Tráns. H-D	1 ^{er} pico de C_v	2 ^o pico de C_v
0.00	0.601	—	—	0.601	—
0.25	0.597	—	—	0.598	—
0.50	0.592	—	—	0.590	—
0.75	—	0.575	0.596	0.578	—
1.00	—	0.546	0.625	0.548	0.650
1.10	—	0.527	0.686	0.528	0.660
1.20	—	0.493	0.720	0.500	0.670
1.30	—	0.348	0.795	0.515	0.730
1.40	—	—	0.798	0.520	0.750

Tabla 3.1: **Temperaturas de transición obtenidas del diagrama de poblaciones.**

de la curva de capacidad calorífica coincide siempre con la temperatura de desnaturalización del estado nativo. Para $\epsilon_{ssm} \leq 0.75$, este máximo coincide con el tránsito entre el estado nativo y el desnaturalizado. Por otra parte, cuando $1.00 \leq \epsilon_{ssm} \leq 1.20$, el tránsito que observamos como el pico más intenso en la curva de capacidad calorífica coincide con el tránsito entre el estado nativo y el estado hinchado. Este comportamiento deja de ser cierto para $\epsilon_{ssm} > 1.20$, donde nos encontramos en un régimen de altas presiones. Sin embargo, el segundo tránsito que observamos cuando $\epsilon_{ssm} \geq 1.00$ sólo coincide con el segundo pico de la curva de capacidad calorífica cuando $\epsilon_{ssm} = 1.00$, para el resto de simulaciones existe una diferencia mayor entre ellas. Esto de nuevo corrobora la posibilidad de que el tránsito entre el estado hinchado y desnaturalizado se dé dentro de un intervalo amplio de temperaturas, y no a una temperatura característica, que es lo que observamos en el caso de la primera transición.

Una vez obtenidas las temperaturas de transición que se reflejan en la Tabla 3.1, nos planteamos la posibilidad de intentar reproducir el diagrama de fases p - T para nuestra proteína. En nuestro caso no tenemos medidas directas de la presión, pero sí tenemos medidas con respecto al parámetro ϵ_{ssm} , que de forma cualitativa representa una presión hidrostática. Por tanto, nos planteamos la representación de un diagrama de fases ϵ_{ssm} - T . La Figura 3.18 muestra el diagrama de fases que hemos obtenido, al representar ϵ_{ssm} frente a las diferentes temperaturas de transición que hemos obtenido

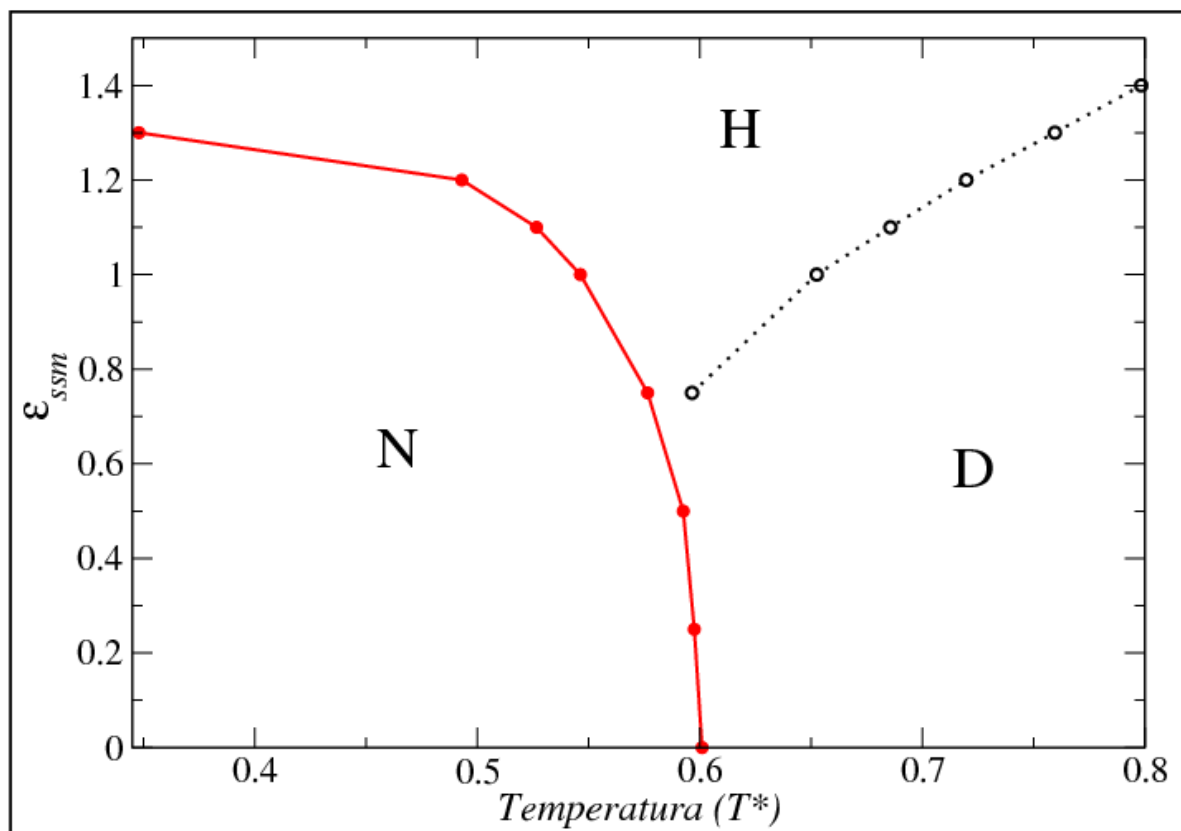


Figura 3.18: **Diagrama de fases $\epsilon_{ssm} - T$ con nuestro modelo.** Representación de ϵ_{ssm} frente a la temperatura. Se indican las temperaturas de transición que hemos obtenido del diagrama de poblaciones. La línea continua roja muestra la transición entre el estado nativo (N) y los estados no nativos. La línea discontinua negra muestra la temperatura aproximada del centro de la transición entre los dos estados no nativos diferentes, el estado hinchado (H) y el estado desnaturalizado (D).

de los diagramas de poblaciones.

El diagrama de fases presenta dos líneas diferentes. La línea continua de color rojo muestra el equilibrio existente entre el estado nativo (N) y los estados no nativos, ya sea el estado desnaturalizado (D) o el estado hinchado (H). Por tanto, esta línea la obtenemos de representar la temperatura del primer tránsito que encontramos en los diagramas de poblaciones. La línea de color negro y discontinua representa el tránsito entre los dos estados no nativos, estado hinchado y estado desnaturalizado. El representarlo como una línea discontinua no es casualidad, puesto que como comentamos en la Sección 1.3.1, una línea continua en un diagrama de fases representa la frontera entre dos estados diferentes del sistema. En nuestro caso el equilibrio entre el estado nativo ('estado cristalino') y desnaturalizado ('estado fundido') es el que observamos con la

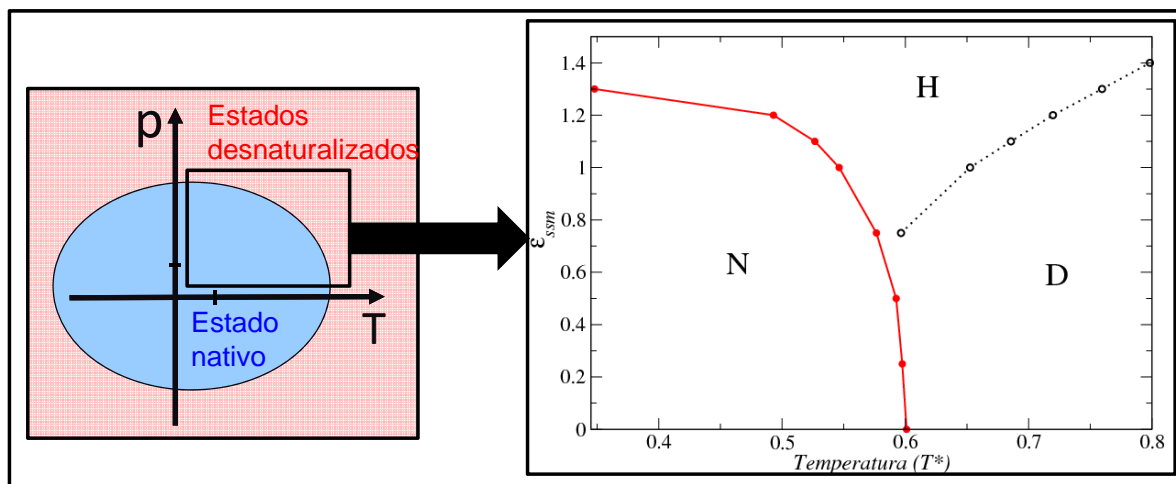


Figura 3.19: **Comparación de nuestro diagrama de fases con el bibliográfico.** La línea que marca la estabilidad del estado nativo en nuestras simulaciones corresponde a una región de la elipse del diagrama de fases descrito en la literatura.

línea continua, ya sea un estado desnaturalizado por efecto de la temperatura o por efecto de la presión. Sin embargo, el tránsito entre dos estados no nativos diferentes no se trata de un equilibrio entre dos estados diferentes sino una transición continua entre diferentes estados ‘líquidos’. Este tránsito se da, como hemos comprobado, de forma continua, y aunque hemos obtenido una temperatura única para caracterizarlo la realidad es que se trata de un proceso continuo que se da en un rango más o menos amplio de temperaturas. Por tanto, la línea discontinua de la Figura representa una temperatura aproximada de transición entre dos tipos de estados desnaturalizados.

Este diagrama de fases que hemos obtenido recuerda mucho al diagrama de fases p - T genérico que presentamos en la Introducción (Figura 1.6), al menos en la zona de presiones y temperaturas superiores a las ambientales (ver Figura 3.19). Por tanto, realmente a través de nuestro modelo hemos sido capaces de reproducir la parte accesible a nuestro modelo del diagrama de fases p - T de las proteínas, al menos para la proteína 2GB1. Se trata de un diagrama de fases cualitativo, pero que al fin y al cabo es capaz de reproducir un estado nativo rodeado de diferentes estados desnaturalizados. Además, en nuestro diagrama trazamos una región de transición entre los estados desnaturalizados por presión y los estados desnaturalizados por efecto de la temperatura, de tal modo

que a la vista de este diagrama podemos saber qué tipo de estado desnaturalizado obtendremos en función de las condiciones por las que abandonemos el estado nativo.

3.7. Análisis a nivel local de las diferentes conformaciones

Una vez comprendido el efecto de la presión en la desnaturalización de la proteína 2GB1, y cómo es posible diferenciar entre dos tipos de estados desnaturalizados, nos planteamos un estudio a nivel local de las diferentes conformaciones que se han ido obteniendo a través de la simulación. De esta manera pretendemos comparar las características estructurales detalladas de los tres estados, dándole especial importancia al estado hinchado, que al fin y al cabo es el estado resultante de un aumento de la presión. Para un estudio detallado de las conformaciones obtenidas es necesario desarrollar algún tipo de herramienta que permita analizar de forma automatizada la multitud de conformaciones que hemos obtenido a lo largo de la simulación para cada valor de ϵ_{ssm} . Recordamos que pese a registrar únicamente un porcentaje pequeño del total de conformaciones generadas, para cada una de las simulaciones acumulamos un mínimo de 2500 fotogramas por temperatura de trabajo, que produce un total de aproximadamente 150000 fotogramas por simulación. Esto hace que sea totalmente imposible realizar un análisis manual de las configuraciones una a una.

Para el análisis de todas estas conformaciones hemos utilizado los denominados **mapas de frecuencia de contactos**. Mediante este tipo de representación obtenemos una información promedio de todas las conformaciones consideradas en el análisis. Esta herramienta consiste en un análisis de cada conformación individual, señalando si cada uno de sus contactos nativos están formados o no, y si lo hacen como contacto directo o como contacto hinchado. El criterio para saber si un contacto se encuentra o no formado en cada configuración es el mismo que se explicó en la Sección 2.4. Una vez

analizado un número dado de configuraciones, el resultado de este análisis se presenta como un mapa de contactos, tal y como los que hemos estado viendo hasta el momento, en el que cada punto en el mapa presentará un color característico, que será función del porcentaje de las conformaciones analizadas que tuviesen ese contacto formado. De esta forma obtendremos un mapa de contactos coloreado en función del porcentaje de veces que los contactos hayan estado formados con relación al número total de configuraciones del análisis. Además, utilizaremos de nuevo mapas de contactos con dos triángulos diferenciados, en el triángulo superior-izquierdo se presenta la frecuencia de los contactos directos, mientras que en el triángulo inferior-derecho se refleja la frecuencia de los contactos hinchados.

Con este tipo de análisis debemos tener un control especial en el tipo de configuraciones que queremos incluir en el análisis, puesto que obtenemos un resultado promedio. Por eso, debemos analizar siempre conformaciones razonablemente similares, que pertenezcan al mismo estado, para que así los resultados tengan algún tipo de interés, ya que si por ejemplo analizamos conformaciones del estado nativo y del estado desnaturalizado conjuntamente, los resultados que obtengamos en nuestro mapa de frecuencia de contactos coincidirán con una conformación intermedia entre ambos estados, pero no se tratará de una conformación real, sino de una conformación creada a partir de otras dos. Por el contrario, si agrupamos conformaciones pertenecientes al mismo estado, y analizamos por tanto conformaciones similares, el mapa de frecuencia de contactos que obtengamos será un mapa promedio que contenga la información característica del estado que estamos estudiando en ese momento. Por tanto, antes de realizar ningún tipo de análisis hay que hacer una clasificación de los distintos fotogramas y agruparlos en estados similares.

En nuestro caso decidimos realizar el estudio clasificando las conformaciones obtenidas de la simulación en los tres estados que venimos definiendo a lo largo de este Capítulo, con lo que la clasificación de cada fotograma se ha hecho bajo el mismo criterio

que habíamos utilizado para los estudios de poblaciones. Serán los valores de radio de giro reducido y RMSD los que dictaminen a cuál de los tres estados, nativo, hinchado o desnaturalizado pertenece cada una de las configuraciones obtenidas a lo largo del muestreo del sistema. En algunos casos, encontramos una alternativa para no tener que clasificar los fotogramas en función de sus valores de radio de giro reducido y RMSD en el diagrama de poblaciones de la Figura 3.17. Si utilizamos conformaciones obtenidas en condiciones de presión y temperatura donde únicamente es estable uno de los tres estados, nos aseguramos de que el mapa de frecuencia de contactos que obtenemos considerando todos los fotogramas sea representativo del estado en cuestión. De esta manera conseguimos un doble objetivo. Por un lado aumentar el número de conformaciones sobre los que hacer la estadística, lo cual siempre ayuda a obtener unos resultados más significativos. Pero además evitamos el análisis de las configuraciones más cercanas a los límites establecidos de radio de giro reducido y RMSD, logrando una menor dispersión de las configuraciones estudiadas, lo que nos lleva de nuevo a unos resultados con menor error.

Existen además algunas consideraciones de carácter general que debemos tener en cuenta cuando queremos analizar los mapas de frecuencia de contactos, que conviene mencionar antes de empezar a mostrar los resultados obtenidos:

- **Sólo los contactos presentes en el mapa de contactos nativos estarán presentes en el mapa de frecuencia de contactos.** Los contactos que aparecen en los mapas de frecuencia son exactamente los mismos que aparecen en su mapa de contactos nativos, no se considera ningún nuevo contacto. Sólo en el caso que la frecuencia de un contacto nativo fuese del 0 %, y por tanto que ninguna conformación analizada tuviese ese contacto formado, el contacto no aparecería en el mapa de frecuencia de contactos.
- **La suma de los valores de frecuencia en los dos triángulos para un mismo contacto nunca puede ser superior al 100 %.** Un contacto determinado

puede encontrarse como contacto nativo o como contacto hinchado, pero nunca por motivos geométricos podrá encontrarse simultáneamente en ambas diagonales en una conformación dada.

- **La suma de la frecuencia de ambas diagonales para un contacto determinado no tiene por qué sumar el 100 %.** Como explicamos detalladamente en la Sección 2.4, consideramos que un contacto nativos se encuentra formado, ya sea como contacto directo o como contacto hinchado, cuando las distancias entre los aminoácidos cumplen ciertas condiciones. Por eso, en los casos en que las distancias entre estos aminoácidos son diferentes se considera el contacto no formado, y no computa en el cálculo de la frecuencia de aparición. Por tanto, la suma del triángulo superior más el triángulo inferior más los casos de contacto no formado es la que da como resultado una frecuencia del 100 %.

3.7.1. El estado nativo y el estado desnaturalizado

Para poder hacer un buen estudio del estado hinchado, el más interesante dentro de los objetivos de este trabajo, parece necesario entender primero los otros dos estados presentes en el proceso de plegamiento. En la Figura 3.20 mostramos los mapas de frecuencias de contactos característicos del estado nativo (a) y del estado desnaturalizado (b). Los mapas de frecuencia de contactos han sido calculados a partir de los resultados de las simulaciones para $\epsilon_{ssm} = 0.00$, donde a distintas temperaturas se estabilizan ambos estados. Para el estado nativo hemos seleccionado una temperatura por debajo de T_m , $T^* = 0.500$, donde el estado nativo es lógicamente el único estable. Para el estado desnaturalizado hemos seleccionado una temperatura por encima de T_m , $T^* = 0.700$.

Analizando este mapa, observamos cómo para el estado nativo el mapa de frecuencias es prácticamente igual al mapa de contactos nativos de la proteína. El triángulo superior donde se muestran los contactos nativos directos es idéntico al de la estructura nativa (Figura 3.2). No sólo están formados todos los contactos nativos, sino que su

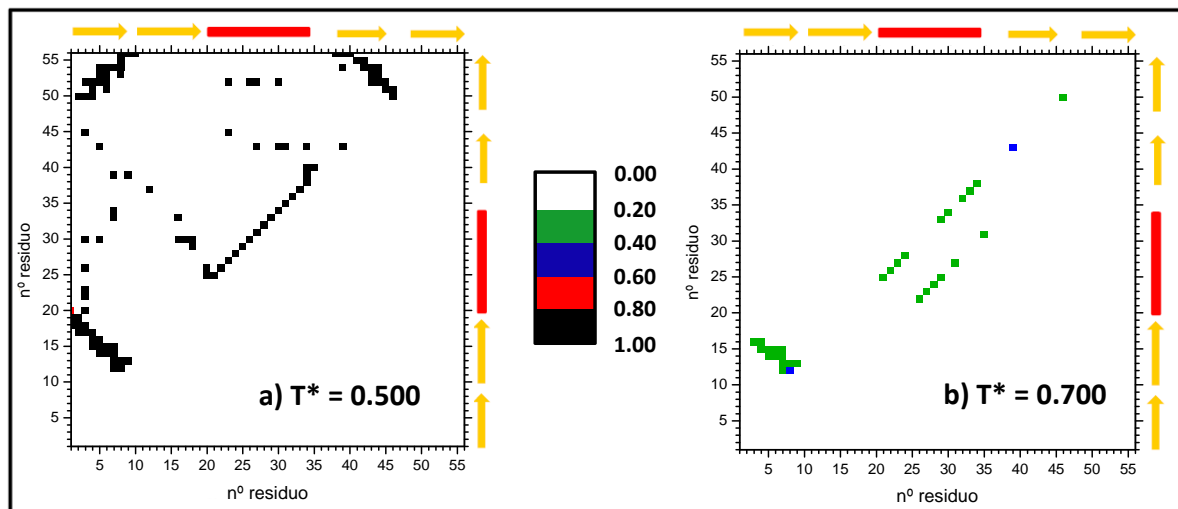


Figura 3.20: **Mapas de frecuencia de contactos para los estados nativo y desnaturalizado.** a) Frecuencia de contactos del estado nativo para $T^* = 0.500$. b) Frecuencia de contactos del estado desnaturalizado para $T^* = 0.700$. Ambos mapas se presentan para un valor de $\epsilon_{ssm} = 0.00$.

frecuencia es siempre cercana a la unidad, lo que supone que cerca del 100 % de los contactos nativos están formado. Como cabía esperar, el triángulo inferior esta completamente vacío puesto que este estado no estabiliza ningún contacto hinchado. Por tanto, nuestro modelo caracteriza correctamente el estado nativo de la proteína a presión atmosférica, como ya habíamos comprobado.

Más interesante resulta el análisis del estado desnaturalizado en la Figura 3.20 (b). En este caso podemos observar en el mapa cómo lo que se obtenemos es un estado sin estructura definida, apenas encontramos contactos formados con población significativa ya sea en el triángulo superior o en el inferior. Sí que cabe destacar que los pocos contactos formados, ya sean contactos directos o hinchados, son todos contactos que estabilizan los elementos de estructura secundaria, apareciendo en las horquillas β o en la hélice α . Sin embargo, apenas se observan los contactos que sirven para unir los diferentes elementos de estructura secundaria y que por tanto estabilizan la estructura terciaria (elipses de color azul en la Figura 3.2). En cualquier caso, esto sólo indica que esporádicamente algunas de las configuraciones mantienen parte de su estructura secundaria formada, aunque la proteína no mantenga ningún tipo de estructura global.

A la vista de estos mapas, y completándolos con un análisis visual de algunos de los fotogramas que hemos ido almacenando a lo largo de la simulación, podemos describir a bajas presiones un estado nativo perfectamente formado, por tanto una proteína globular con los elementos de estructura secundaria bien organizados y con baja fluctuación estructural, a la vista de que la frecuencia de todos sus contactos es superior al 80 %. Por otro lado se describe un estado desnaturalizado muy poco estructurado, como una cadena abierta de aminoácidos, donde sólo algunos de ellos forman contactos esporádicos. Además, la poca frecuencia presente en cada contacto es característica de grandes fluctuaciones conformacionales en el sistema, puesto que las diferentes configuraciones presentes en este estado son muy diferentes entre sí. Cada una de las conformaciones presentes tiene formados sólo unos pocos contactos nativos, distintos de unas a otras, y el promedio sobre todas ellas es la que nos da una frecuencia tan baja. Por tanto, al menos a presión atmosférica, el modelo es capaz de reproducir las estructuras características tanto de la estructura nativa de la proteína, cuando nos encontramos por debajo de T_m , como las conformaciones del estado desnaturalizado de la proteína cuando nos encontramos por encima de T_m .

3.7.2. Comportamiento estructural a alta presión

Una vez comprendido el mapa de frecuencias de contactos característico de los dos estados ya conocidos, nos planteamos realizar un análisis de las diferentes configuraciones obtenidas a lo largo de las simulaciones a presión moderada. En los mapas de poblaciones ya vimos cómo para valores medios de ϵ_{ssm} podemos aislar los tres estados diferentes en función de la temperatura. En la Figura 3.21 se muestran los mapas de frecuencias de contactos a cuatro temperaturas características cuando $\epsilon_{ssm} = 1.20$, el valor más alto de ϵ_{ssm} dentro del régimen de presiones moderadas. Las cuatro temperaturas seleccionadas, según nuestro diagrama de fases ϵ_{ssm} - T , corresponden al estado nativo cuando $T^* = 0.450$ (a), al estado hinchado cuando $T^* = 0.530$ y 0.650 , (b) y (c), y al

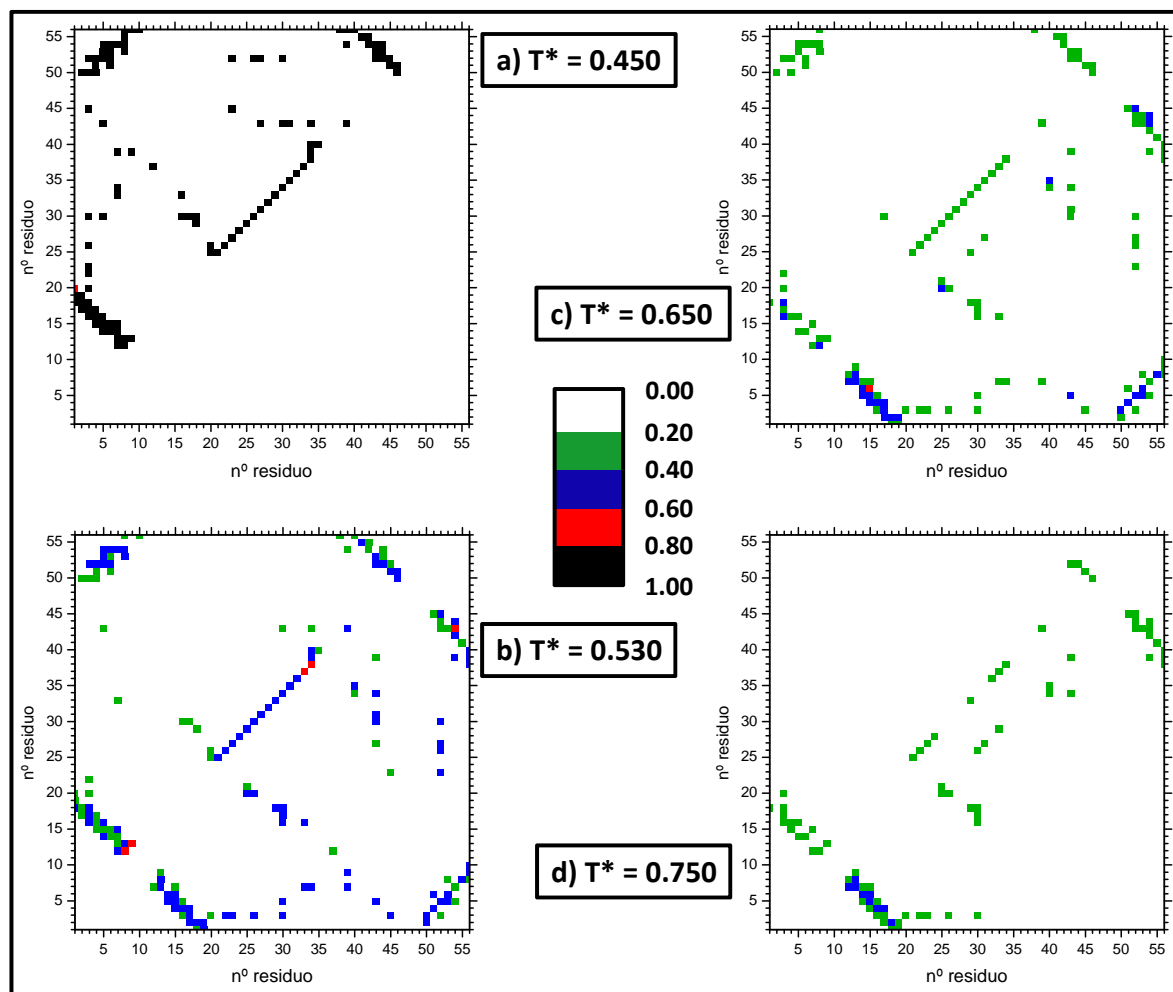


Figura 3.21: **Evolución de la estructura con la temperatura a presión moderada.** Mostramos los mapas de frecuencia de contactos que hemos obtenido en condiciones de presión moderada ($\epsilon_{ssm} = 1.20$), a diferentes temperaturas. Cada uno de los mapas muestra uno de los estados característicos de este tránsito: a) Estado nativo, $T^* = 0.450$. b) Estado hinchado, $T^* = 0.530$. c) Estado hinchado, $T^* = 0.650$. d) Estado desnaturalizado, $T^* = 0.750$.

estado desnaturalizado cuando $T^* = 0.750$ (d). De esta manera, analizando los mapas podremos ver, aunque sea de forma discreta, la evolución de la estructura a medida que aumenta la temperatura dentro de unas condiciones de alta presión.

Analizando los mapas de frecuencia de contactos podemos apreciar una notable evolución de estos con la temperatura. Cuando $T^* = 0.450$ nos encontramos con el mapa propio del estado nativo, prácticamente idéntico al que habíamos mostrado para el estado nativo a presión atmosférica en la Figura 3.20 (a). Observamos cómo a pesar de encontrarnos en condiciones de presión moderada el estado nativo está totalmente

estabilizado.

Cuando la temperatura aumenta, y por tanto según nuestro diagrama de fases nos encontramos en el régimen del estado hinchado, podemos apreciar un cambio significativo en la estructura de la proteína. Así, para $T^* = 0.530$ obtenemos un estado muy diferente tanto del estado nativo como del estado desnaturalizado que habíamos mostrado previamente a presión atmosférica. A esta temperatura apreciamos cómo muchos de los contactos nativos directos se han perdido, y a cambio se han formado muchos contactos hinchados. Analizando en el mapa cuáles han sido los contactos que se han visto más modificados, apreciamos cómo para la mayoría de los contactos que estabilizan la estructura terciaria la frecuencia como contacto hinchado es mayor a la frecuencia como contacto directo. Un aumento de 3 Å en la distancia entre contactos que estabilizan la estructura terciaria corresponde a un aumento del volumen en el interior de la proteína, ya que la distancia existente entre los diferentes elementos de estructura secundaria se hace mayor de la que tenía la proteína en el estado nativo. Además, observamos cómo los contactos que estabilizan la formación de elementos de estructura secundaria no se encuentran tan afectados por el segundo pozo del potencial y su frecuencia queda más compensada entre ambos mínimos del potencial. Especialmente interesante es ver las diferencias entre los dos elementos diferentes de estructura secundaria, ya que la hélice α está bastante menos influida por el aumento de presión que la lámina β . La discusión de cómo afecta el aumento de la presión a los diferentes elementos de estructura secundaria lo abordaremos en detalle a lo largo del siguiente Capítulo. Sin embargo, si vemos cómo gracias en parte a los contactos de enlace de hidrógeno que introducimos en el mapa de contactos, la estructura secundaria está parcialmente formada, como es característico de los estados desnaturalizados por presión⁴⁹. Por tanto, parece que el modelo de simulación es capaz de reproducir las características de la estructura desnaturalizada por presión, que inspiraron nuestra definición del potencial de interacción.

La siguiente temperatura que mostramos, $T^* = 0.650$, corresponde de nuevo al régimen del estado hinchado. Observamos cómo a pesar de encontrarnos de nuevo en el mismo estado, la configuración promedio que se muestra ha evolucionado considerablemente, se han perdido muchos de los contactos formados que observábamos para $T^* = 0.530$. La pérdida de contactos se produce tanto en el triángulo superior como en el triángulo inferior, lo cual es característico de los estados desnaturalizados por temperatura. Por tanto, parece que nos encontramos con una estructura intermedia entre el estado hinchado y el estado desnaturalizado. Esta situación intermedia entre ambos estados confirma nuestro análisis previo de que entre el estado hinchado y el estado desnaturalizado existe un tránsito continuo y no un equilibrio de dos estados. Por tanto, el estado hinchado va a ir perdiendo su estructura característica a medida que aumenta la temperatura.

Por último, se muestra el mapa de frecuencia de contactos cuando $T^* = 0.750$, Figura 3.21 (d). En este caso nos encontramos en el régimen del estado desnaturalizado, y de hecho obtenemos un mapa característico de este estado. Apenas se encuentran contactos formados, y los pocos que lo están son contactos que estabilizan la estructura secundaria, mientras que los contactos que estabilizan la estructura terciaria han desaparecido completamente. Por tanto, cuando la temperatura es suficientemente elevada, se muestra cómo la estructura desnaturalizada por presión evoluciona perdiendo incluso los contactos hinchados formados hasta que se obtiene la estructura desnaturalizada.

De esta manera, hemos comprobado cómo a una presión suficientemente alta obtenemos tres estados diferentes en función de la temperatura a la que nos encontremos. Además, gracias a los mapas de frecuencia de contactos podemos hacernos una idea de cómo es la estructura del estado hinchado que obtenemos a partir de nuestro modelo. Se trata de una estructura menos compacta que la estructura nativa, ya que entre sus elementos de estructura secundaria existe una mayor distancia, generando un volumen aparentemente vacío en el interior de la proteína, que en realidad estaría ocupado

por el disolvente. Además, observamos cómo los elementos de estructura secundaria se encuentran bastante bien formados pese a tratarse de una estructura desnaturalizada.

3.7.3. El estado hinchado

En la evolución que hemos mostrado del estado de la proteína ha quedado demostrado que existen diferencias significativas entre las estructuras desnaturalizada e hinchada. Al mismo tiempo, como la transición entre ambos estados se trata de un proceso continuo y no cooperativo, obtenemos diferentes configuraciones pertenecientes al estado hinchado que evolucionan en función de la temperatura y presión a la que nos encontremos. Además, puesto que la estabilización del estado hinchado cuando $\epsilon_{ssm} = 1.20$ se da a una temperatura intermedia, sabemos que los mapas de frecuencia de contactos mostrados anteriormente para el estado hinchado presentan bastante fluctuación térmica. Por esa razón decidimos estudiar la estructura hinchada a una temperatura suficientemente baja como para que se minimicen las fluctuaciones térmicas del sistema. Así comprendemos cuál es el estado hinchado de referencia al que nos lleva el modelo desarrollado. En la Figura 3.22 se muestra el mapa de frecuencia de contactos obtenido a la temperatura más baja de estudio, $T^* = 0.300$, cuando $\epsilon_{ssm} = 1.40$. Para este valor de ϵ_{ssm} , como se puede ver en el diagrama de fases ϵ_{ssm} - T (Figura 3.18), incluso a la temperatura más baja todas las configuraciones obtenidas corresponden al estado hinchado. Además, tratándose de unas condiciones de alta presión y baja temperatura, el estado hinchado debería estar muy bien caracterizado, con muy pocas fluctuaciones térmicas.

Este mapa de frecuencias por tanto representa un estado hinchado estabilizado en unas condiciones relativamente extremas, a muy altas presiones y temperaturas muy bajas. Nos muestra cómo el modelo de interacción que hemos desarrollado es capaz de reproducir satisfactoriamente los diferentes efectos que se describen experimentalmente para los estados desnaturalizados por presión. Podemos observar cómo se muestra un

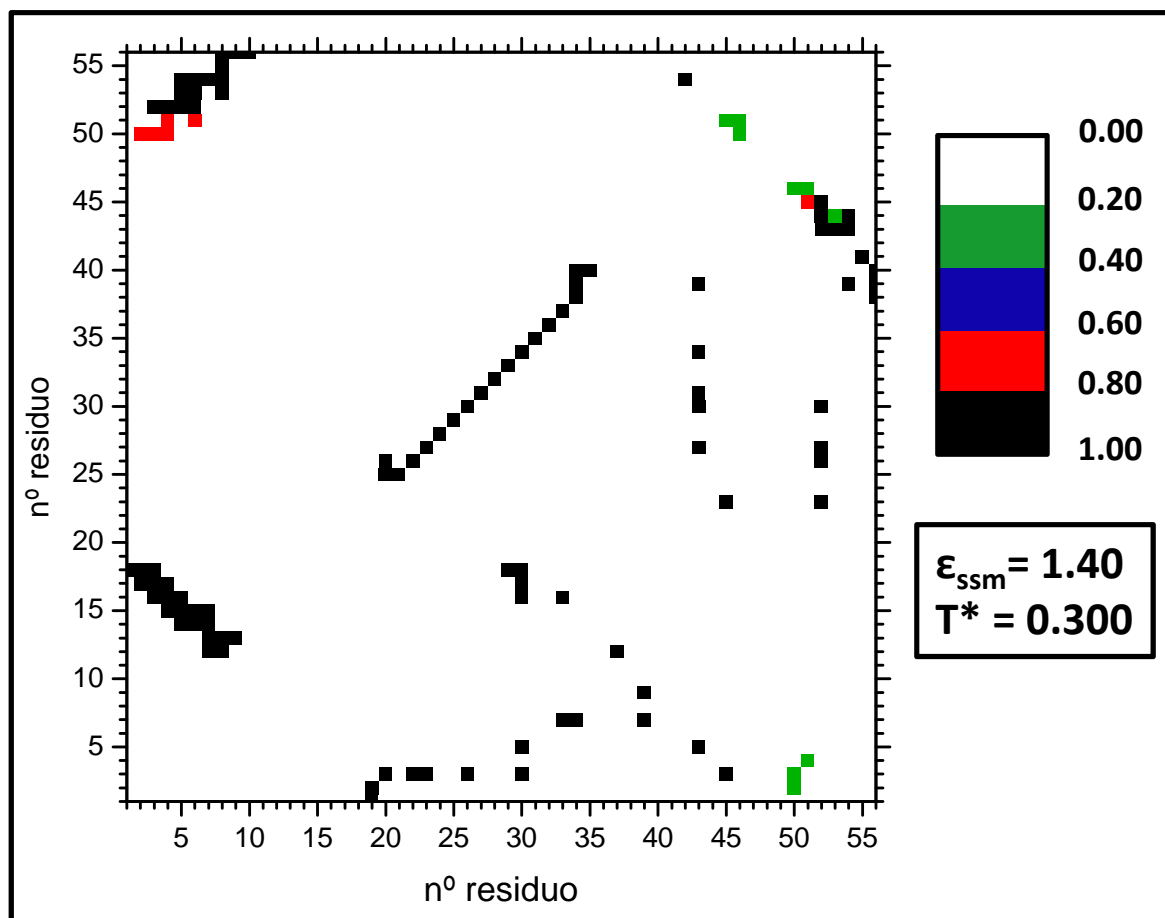


Figura 3.22: **Mapa de frecuencia de contactos para el estado hinchado en condiciones de alta presión.** Se muestra el mapa de frecuencia de contactos obtenido en unas condiciones relativamente extremas de alta presión ($\epsilon_{ssm} = 1.40$) y muy bajas temperatura ($T^* = 0.300$).

estado hinchado con la hélice correctamente formada, ya que todos los contactos que la estabilizan se encuentran como contactos nativos directos. Además, tres de las cuatro hebras β también mantienen todos sus contactos dentro del pozo Gō. Por otro lado, los contactos que estabilizan la estructura terciaria se encuentran en su totalidad como contactos hinchados y además en una frecuencia cercana al 100%. Por tanto, en este diagrama se muestra el estado hinchado por excelencia al que nos llevaría el modelo en caso de obtener un proceso de desnaturalización únicamente dirigido por la presión y no por el efecto conjunto de presión y temperatura.

3.7.4. Evolución del estado hinchado con la presión.

Una vez mostrados los mapas característicos de los tres estados presentes en nuestras simulaciones, nos planteamos un último estudio de las configuraciones obtenidas a lo largo de la simulación. Puesto que somos conscientes de que existe una evolución del estado hinchado en función de la presión del sistema, realizamos una comparación de los diferentes estados hinchados obtenidos en simulaciones a diferentes presiones. Así, en la Figura 3.23 se muestran los mapas de frecuencias de contactos obtenidos para diferentes valores de ϵ_{ssm} . Para que la comparación entre configuraciones procedentes de diferentes simulaciones pueda ser realizada de forma sencilla, los mapas de frecuencias de contacto están realizados a una misma temperatura para todos los valores de ϵ_{ssm} , $T^* = 0.570$. Para esta temperatura, en las simulaciones donde $\epsilon_{ssm} \geq 1.20$, el estado hinchado es siempre el único estado poblado, como se puede ver en los mapas de poblaciones correspondientes (Figura 3.17). Utilizar siempre la misma temperatura nos permite comparar a través de los diferentes mapas el efecto del aumento de la presión eliminando, al menos en parte, el efecto de la temperatura, puesto que a una temperatura determinada todas las configuraciones presentaran fluctuaciones térmicas similares. Los cuatro mapas de frecuencia de contactos presentes en la Figura 3.23, reflejan la presencia del estado hinchado. Sin embargo, se aprecian diferencias entre las características de este estado en función de la presión a la que ha sido sometido el sistema. Notamos que cuando la presión es moderada, $\epsilon_{ssm} = 1.10$, panel (a), el estado hinchado se encuentra parcialmente estabilizado. A medida que aumenta el valor de ϵ_{ssm} se observa cómo parte de sus contactos nativos se transfieren desde el pozo Gō hasta el pozo ssm. Además, los contactos que sirven para estabilizar los elementos de estructura terciaria son cada vez más frecuentes como contactos hinchados a medida que aumenta la presión. En cuanto a los elementos de estructura secundaria presentes en la estructura, la hélice es bastante estable frente al aumento de la presión, mientras que los contactos que caracterizan a la lámina β se ven más afectados. Cuando $\epsilon_{ssm} \geq 1.3$, observamos cómo los contactos que

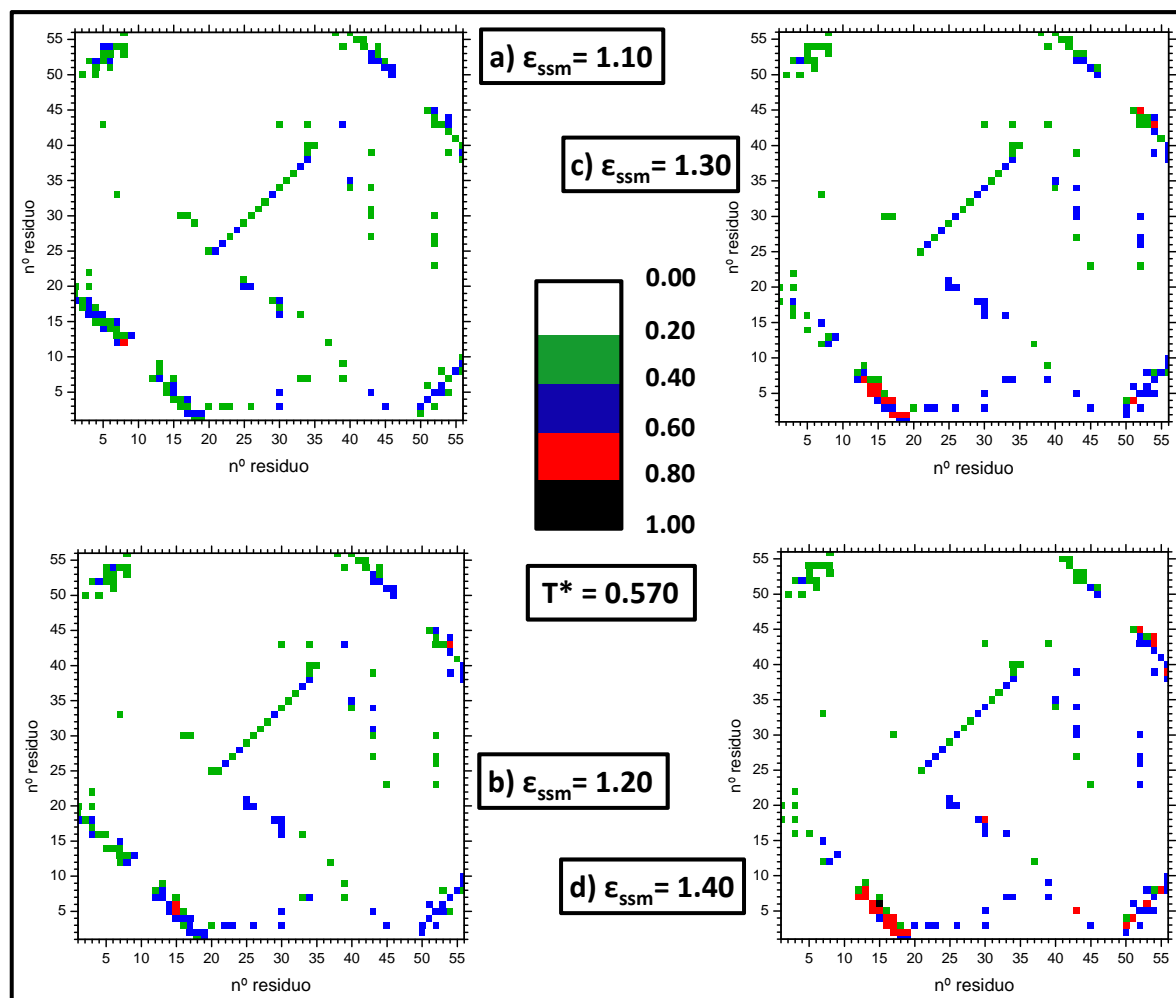


Figura 3.23: **Evolución del estado hinchado con la presión.** Evolución de los mapas de frecuencia de contactos para el estado hinchado en función del valor de ϵ_{ssm} . Todos los mapas que se presentan están realizados para $T^* = 0.570$. a) $\epsilon_{ssm} = 1.10$. b) $\epsilon_{ssm} = 1.20$. c) $\epsilon_{ssm} = 1.30$. d) $\epsilon_{ssm} = 1.40$.

estabilizan a la lamina β se encuentran en su mayoría como contactos hinchados. Por tanto, vemos cómo a medida que aumenta la presión el estado hinchado de la proteína también se ve afectado, y va evolucionando hacia una conformaciones con estructura menos definida.

Con esta última discusión del estado hinchado en función de la presión hemos analizado las conformaciones más interesantes que aparecen a lo largo de la simulación, observando la diferencia existente entre los tres estados, nativo, hinchado y desnaturizado, entendiendo las principales diferencias entre los mapas de frecuencia de contactos que caracterizan a cada estado. Hemos analizado también las diferentes configuraciones

obtenidas a lo largo de una simulación de alta presión, y hemos analizado la evolución que se observa en el estado hinchado a medida que se transforma en el estado desnaturalizado por un aumento de la temperatura. También hemos observado los cambios que sufre el estado hinchado a medida que aumenta la presión. Todos estos análisis han servido para que podamos conocer en detalle las características estructurales que presenta la proteína a cada valor de presión y temperatura.

3.8. Resumen y conclusiones del Capítulo

En este primer Capítulo de resultados hemos estudiado el proceso de plegamiento/desnaturalización, debido los efectos conjuntos de la presión y temperatura, del dominio B1 de la proteína G (2GB1, según su código PDB). La primera parte del Capítulo está dedicada a introducir la proteína seleccionada, específicamente su estructura y el mapa de contactos que corresponde a esta proteína para que pueda ser simulada con nuestro modelo. Esta primera parte del Capítulo también incluye un estudio de los tres parámetros de nuestro potencial de interacción, que nos ha permitido encontrar una relación cualitativa entre el valor de ϵ_{ssm} y la presión del sistema. A continuación hemos explicado la metodología que aplicamos para poder reproducir un aumento de la presión hidrostática en nuestro modelo de simulación. Esta metodología consiste en realizar diferentes simulaciones haciendo crecer el valor de ϵ_{ssm} , puesto que un valor mayor de ϵ_{ssm} reproduce en un esquema de campo medio un aumento de la presión hidrostática del medio. Las simulaciones realizadas para diferentes valores de ϵ_{ssm} corresponden entonces a sistemas a diferentes presiones. Como además, para cada valor de ϵ_{ssm} se han simulado diferentes temperaturas de trabajo, obtenemos resultados para nuestra proteína dentro de un intervalo amplio de temperaturas y de presiones. Por último, antes de mostrar los resultados de la simulación para esta proteína, hemos mostramos algunas de las herramientas de análisis que vamos a utilizar para poder estudiar el tipo de plegamiento que presenta esta proteína.

En la segunda parte de este Capítulo hemos mostrado los resultados que hemos obtenido en la simulación de la proteína 2GB1 al aplicar nuestro modelo de altas presiones tal y como lo hemos descrito en la primera parte del Capítulo. En primer lugar, hemos mostrado las diferentes curvas de capacidad calorífica que hemos obtenido para las nueve presiones utilizadas en esta proteína, $\epsilon_{ssm} = 0.00, 0.25, 0.50, 0.75, 1.00, 1.10, 1.20, 1.30$ y 1.40 . A partir del análisis de las curvas de capacidad calorífica, junto con el análisis de los perfiles de energía de Helmholtz de los diferentes tránsitos, identificamos tres regímenes bien diferenciados de presión en función del valor de ϵ_{ssm} . Así, cuando $\epsilon_{ssm} < 0.75$ nos encontramos en un régimen de baja presión, cuando ϵ_{ssm} está entre 0.75 y 1.2 hablaremos de presiones moderadas, mientras que cuando $\epsilon_{ssm} > 1.2$ hablamos de un régimen de altas presiones. Posteriormente, gracias a los histogramas mono y bidimensionales de dos propiedades estructurales como son el radio de giro y RMSD, confirmamos la presencia de un estado desnaturalizado por presión que difiere estructuralmente del estado desnaturalizado ‘clásico’ debido a la temperatura. Por tanto, en el proceso de plegamiento/desnaturalización además de los estados nativo (N) y desnaturalizado (D), existe un estado no nativo al que denominamos estado hinchado (H). Mediante el análisis de los histogramas confirmamos la existencia de diferentes tipos de tránsitos entre los estados involucrados. Así contemplamos un equilibrio cooperativo de dos estados entre N y D a bajas presiones; un equilibrio de dos estados entre los estados N e H que se va haciendo menos cooperativo conforme aumenta la presión del sistema; y por último un tránsito no cooperativo entre los estados H y D. También gracias a los histogramas podemos seleccionar unos valores de las dos propiedades estructurales, radio de giro reducido y RMSD, que nos permiten clasificar cada una de las conformaciones obtenidas en uno de los tres estados. Esto a su vez nos permite realizar un estudio cuantitativo de las poblaciones obtenidas para cada uno de los estados en función de la temperatura. Cuando realizamos el análisis de poblaciones para todas las presiones, obtenemos hasta dos temperaturas de transición, en función del número de estados y de

los tránsitos involucrados a cada presión. La representación de las temperaturas de transición frente al valor del parámetro ϵ_{ssm} nos permite obtener nuestro propio diagrama de fases ϵ_{ssm} - T , que resulta ser muy similar al que se describe en la bibliografía.

Por último, a través de los mapas de frecuencia de contactos hemos realizado un análisis estructural sistemático de las configuraciones obtenidas en diferentes regiones del espacio p - T . Mediante esta herramienta hemos podido comprender mejor las características de la estructura hinchada de esta proteína, que se caracteriza por mantener aproximadamente los elementos de estructura secundaria bien formados, pero más alejados entre ellos que en el estado nativo. Además, hemos comprobado que al aumentar la temperatura el estado hinchado se transforma de forma gradual en el estado desnaturalizado, mientras que cuando existe un aumento de la presión a temperatura constante el estado hinchado también evoluciona, manteniéndose siempre en estructuras más compactas que las desnaturalizadas térmicamente.

Por último, y a modo de resumen de este Capítulo, en la Figura 3.24 se muestra una representación de ϵ_{ssm} frente a la temperatura (T^*), donde cada punto en la gráfica muestra una de las condiciones de presión y temperatura a la que hemos realizado la simulación. Con esta imagen perseguimos un doble objetivo. Por un lado que quede reflejada la gran cantidad de condiciones en las que hemos explorado el embudo de plegamiento. A la derecha de la imagen, entre paréntesis, se muestra el número de temperaturas que hemos utilizado para cada valor de ϵ_{ssm} . En total suponen alrededor de 450 condiciones presión/temperatura para esta proteína. Pero además aprovechamos la imagen para señalar en distintos colores cada una de las condiciones de simulación que hemos ido analizando a lo largo de este Capítulo.

Así, y según consta en la leyenda de la imagen, los puntos y líneas de color rojo y azul representan las líneas de transición de fases que hemos mostrado en nuestro diagrama de fases ϵ_{ssm} - T (Figura 3.18). Además, hay que aclarar que los puntos marcados en estos dos colores son los únicos que no representan condiciones específicas de

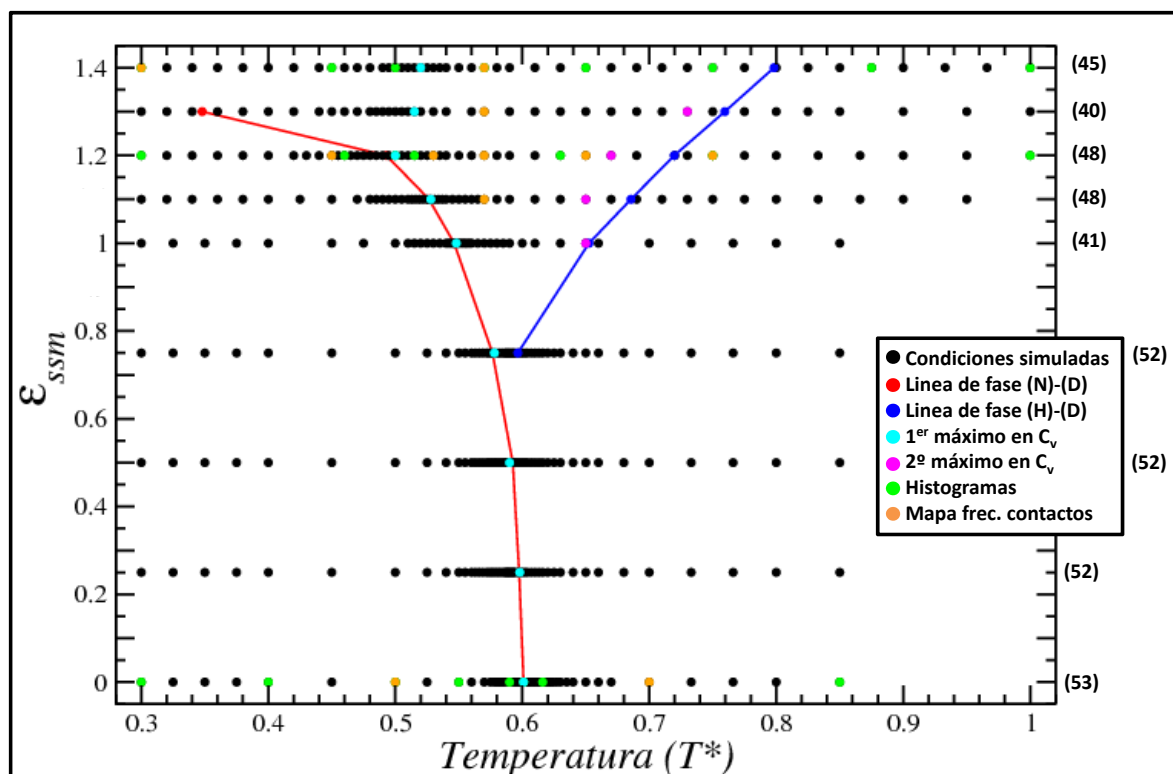


Figura 3.24: **Diagrama de fases indicando los puntos de la simulación.** Resumen de las condiciones de presión y temperatura a las que hemos explorado el espacio conformacional de plegamiento de la proteína 2GB1.

simulación, puesto que como ya comentamos dichos puntos se obtienen de los puntos de corte de las gráficas de poblaciones. En color azul claro se muestran las condiciones a las que se ha obtenido el máximo de la capacidad calorífica para cada valor de ϵ_{ssm} . En rosa se presentan las condiciones a las que se ha obtenido el segundo máximo de la capacidad calorífica para los valores más altos de ϵ_{ssm} . En verde se presentan condiciones de presión y temperatura que han sido seleccionadas para su análisis a lo largo de este Capítulo para estudiar los histogramas mono y bidimensionales de sus radios de giro y RMSD. Por último, en naranja se presentan las condiciones de presión y temperatura a las cuales hemos analizado diferentes mapas de frecuencia de contactos. De esta forma, a través de una única imagen, queremos reflejar todo el trabajo y los diferentes análisis que hemos realizado para poder comprender el proceso de plegamiento de la proteína 2GB1, bajo unas condiciones variables de presión y temperatura.

La conclusión principal a la que podemos llegar tras analizar los diferentes resultados mostrados a lo largo del Capítulo es que el modelo de simulación que hemos desarrollado es capaz de reproducir de una forma adecuada el efecto que ejerce un aumento de la presión hidrostática en el proceso de desnaturalización de esta proteína. Además, hemos apreciado que, para nuestro modelo, el efecto del aumento de la presión empieza a ser apreciable cuando la relación de los parámetros $\epsilon_{ssm}/\epsilon_{go} > 1$. Esta situación es la que cabría esperar a tenor del potencial de interacción que hemos desarrollado, puesto que con esta relación de los parámetros el pozo ssm resulta ser el más profundo y por tanto se deja de estabilizar al estado nativo.

Por último, cabe destacar cómo obtenemos diferentes tipos de tránsitos entre los tres estados involucrados, N, H y D. Además, las características de las diferentes transiciones evolucionan con la presión y la temperatura de una manera muy similar a como se describe en la bibliografía. El hecho de que el modelo de simulación sea capaz de reproducir las características de los distintos tránsitos a pesar de no haber impuesto ningún tipo de restricción para que esto suceda, ratifica nuevamente que el modelo funciona de forma correcta.

Capítulo 4

Evaluación del modelo en proteínas con distinta estructura secundaria

En el modelo de simulación desarrollado en esta Tesis, que hemos presentado y puesto a punto en los dos capítulos anteriores de esta memoria, hemos comprobado que el proceso de desnaturalización de una proteína viene dictaminado por el mapa de contactos de su estructura nativa, que sirve de base para la descripción de las interacciones presentes en el sistema. Hemos comprobado cómo, mediante la aplicación de nuestro modelo, la distribución de los diferentes contactos nativos que aparecen en el mapa como contactos locales, contactos de largo alcance y contactos de enlace de hidrógeno es la que dictamina en buena medida la estructura del estado hinchado, y por lo tanto también las características de la transición que aparece entre los diferentes estados involucrados. En este Capítulo queremos analizar con más detalle esta situación, y específicamente las posibles diferencias que puede inducir nuestro modelos en proteínas con diferentes elementos de estructura secundaria. Además, en la bibliografía se recoge cómo los dos principales elementos de estructura secundaria pueden verse afectados de manera diferente por el aumento de la presión hidrostática^{93,166}.

Observando los mapas de contactos de proteínas diferentes se puede apreciar, co-

mo mostraremos a continuación, que existe una distribución distinta de los contactos nativos entre los diferentes tipos mencionados arriba, cuando se comparan entre sí los dos principales elementos de estructura secundaria, las hélices α y las láminas β . Cada uno de los dos motivos de estructura secundaria se estabiliza de forma mayoritaria gracias a una contribución específica de los diferentes tipos de contactos. No olvidemos que la clasificación de los contactos nativos en contactos de diferente tipo no la realizamos de forma arbitraria, sino que, como ya explicamos en la Sección 2.2.2, es la propia estructura nativa, determinada experimentalmente para una proteína dada, la que define los enlaces de hidrógeno presentes entre diferentes aminoácidos en la estructura nativa de la proteína. Recordemos también que, en nuestro modelo, estos contactos de enlace de hidrógeno, al igual que las interacciones locales, presentan un solo mínimo en el potencial de interacción, correspondiente a lo que hemos denominado ‘contacto directo’. Solo los contactos de largo alcance que no corresponden a enlaces de hidrógeno nativos presentan un segundo mínimo, cuya profundidad depende de la presión ejercida sobre el sistema, que representa la inyección de una molécula de agua entre los residuos que interaccionan. Por lo tanto, la disposición específica de los contactos que estabilizan los elementos de estructura secundaria y la estructura terciaria global, junto con su clasificación como contactos de largo alcance o contactos de enlace de hidrógeno, puede dar lugar a situaciones bastante diferentes. El hecho de que hélices α y láminas β correspondan a situaciones bien diferenciadas en este contexto es lo que pretendemos analizar de forma detallada en este Capítulo.

Poder encontrar la relación existente entre las características del proceso de plegamiento de una proteína y sus elementos de estructura secundaria supone una herramienta esencial para continuar comprendiendo cómo afecta un aumento de la presión del medio a la estabilidad y desnaturalización de una proteína determinada. Al mismo tiempo, el hecho de que nuestro modelo pudiera ser capaz de reproducir los efectos de la presión observados experimentalmente en los dos motivos principales de elementos de estructu-

ra secundaria constituiría una validación definitiva de sus posibilidades. Sin introducir ninguna condición adicional a la simulación que no esté de por sí incluida en el mapa de contactos, los resultados que presentamos en este Capítulo nos permitirán confirmar cómo el modelo que hemos desarrollado resulta idóneo para la simulación en función de la presión de proteínas con diferente estructura secundaria.

4.1. Motivación del Capítulo

A lo largo de este segundo Capítulo de resultados, pretendemos extender nuestro modelo de simulación al proceso de desnaturalización de diferentes proteínas. Enfocaremos nuestro análisis en estudiar cómo el efecto de un aumento de la presión del medio influye en los dos elementos principales de estructura secundaria de las proteínas, la lámina β y la hélice α . Para realizar este estudio utilizaremos cuatro proteínas distintas, pertenecientes a dos familias estructurales diferentes: las proteínas α , que se caracterizan por estabilizar únicamente hélices α en su estructura nativa, y las proteínas β , que estabilizan únicamente estructura β en su estado nativo. Nótese que aunque hablamos de familias estructurales que estabilizan un solo tipo de estructura secundaria, las proteínas de una misma familia no tienen por qué ser similares entre sí, ya que cada una de las proteínas puede estar compuesta por diferente número de elementos de estructura secundaria. Además, los elementos individuales pueden ser de diferente tamaño y se pueden organizar de muchas maneras posibles. Por tanto, a pesar de ser familias compuestas por un único tipo de estructura secundaria, son familias de proteínas con mucha variabilidad estructural. De esta manera pretendemos abordar varios objetivos a lo largo del Capítulo.

Por un lado, queremos continuar estudiando el proceso de desnaturalización de diferentes proteínas en condiciones de alta presión con nuestro modelo de simulación. De esta forma, todos los resultados obtenidos a lo largo del Capítulo se sumarán a los ya presentados para la proteína 2GB1, ampliando notablemente la información acerca de

las posibilidades de nuestro modelo. A través de la simulación de estos nuevos sistemas tendremos que corroborar si el modelo que hemos diseñado es válido para la simulación de proteínas con cualquier estructura, y no sólo para la proteína 2GB1 con la que calibramos nuestro modelo.

Pero sobre todo, los resultados que obtengamos a lo largo de este Capítulo servirán para hacer un estudio de cómo el aumento de la presión afecta a los dos principales elementos de estructura secundaria, al estar estos presentes en las estructuras nativas de buena parte de las proteínas naturales, al menos de aquellas cuya estructura se conoce hasta el momento. Conocer el efecto general de la presión sobre dichos motivos puede resultar esencial para poder entender cómo afecta de forma distinta a proteínas diferentes. Incluso puede resultar muy interesante de cara a ser capaces de predecir, en base a la estructura de la proteína, cómo le afecta un aumento de la presión hidrostática en el medio.

En este punto parece importante mencionar que en los últimos años se han publicado trabajos donde se demuestra que existen diferencias en el comportamiento de los dos principales elementos de estructura secundaria frente a un aumento de presión hidrostática. Así, se ha encontrado que las hélices α son más estables frente al aumento de la presión del medio, y por tanto necesitan de una mayor presión hidrostática para desestabilizar su estructura. Incluso hay estudios que muestran cómo un aumento de la presión, dentro de un intervalo concreto de valores, puede producir un aumento en la estabilidad de las hélices dentro de la estructura proteica^{167,168}. Las láminas β , sin embargo, sufren en mayor medida el aumento de la presión, y se desestabilizan a presiones inferiores a las que lo hacen las hélices α , perdiendo la estructura nativa tipo β que presentan a presión atmosférica^{93,166}. Comprobar si este comportamiento se reproduce con nuestro algoritmo de simulación constituye una forma de seguir comprobando el correcto funcionamiento del modelo desarrollado en esta Tesis.

Por tanto, a lo largo del Capítulo introduciremos en primer lugar los sistemas

proteicos que vamos a analizar, y a continuación iremos mostrando los resultados que hemos obtenido al simular su proceso de desnaturalización bajo presión. El análisis de resultados se centrará en observar los fenómenos que se producen en cada una de las proteínas a medida que aumenta la presión, pero dando especial importancia a cómo evolucionan los estados hinchados de cada una a partir de los diferentes elementos de estructura secundaria que presentan en condiciones nativas. A lo largo de este Capítulo recurriremos en algunos casos a la comparación con los resultados obtenidos para la proteína 2GB1, que al tratarse de una proteína de tipo α/β presenta simultáneamente los dos elementos de estructura secundaria, como ya mostramos en la Figura 3.1.

4.2. Proteínas con un solo tipo de estructura secundaria

Hemos seleccionado cuatro proteínas para simular con nuestro modelo. Todas ellas son proteínas de tamaño relativamente pequeño, y mantienen las características presentes en la proteína 2GB1 y que explicamos en la Sección 3.2 relacionadas con su tamaño, naturaleza de su cadena polipeptídica, etc. Además, todas ellas presentan un plegamiento de dos estados a presión atmosférica. Por tanto, la comparación entre los resultados de todas ellas y la proteína 2GB1 parece más que adecuada.

4.2.1. Proteínas α

Esta familia de proteínas es a priori la más sencilla de definir y clasificar. Consiste en la agrupación de diferentes hélices α dispuestas a lo largo de una única cadena, dando lugar en el plegamiento a una estructura compacta. Como en cualquier proteína globular, la agrupación de las diferentes hélices se estabiliza a través de interacciones favorables entre los residuos hidrófobos que quedan en el interior de la estructura nativa, mientras que los aminoácidos hidrófilos quedan en el exterior de la estructura, interaccionando

con el disolvente.

Hemos seleccionado dos proteínas pertenecientes a esta familia. Manteniendo su nombre tomado del Protein Data Bank, la primera es el *amino-terminal domain of phage 434 repressor*, cuyo código PDB es 1R69^{169,170}, que presenta un plegamiento cooperativo a presión atmosférica¹⁷¹. La otra proteína de esta familia es el *Drosophila engrailed homeodomain*, 1ENH según su código PDB^{172,173}, que según la bibliografía también presenta un plegamiento de dos estados a presión atmosférica⁷⁸. Ambas son proteínas compuestas por cinco y tres hélices α , respectivamente, y por 63 y 54 aminoácidos cada una de ellas. En los paneles (a) y (c) de la Figura 4.1 podemos observar los mapas de contactos de estas dos proteínas, donde mostramos cómo se encuentran distribuidas las diferentes hélices que constituyen cada una de sus estructuras nativas. 1R69 presenta 245 contactos nativos. De ellos, 124 son considerados contactos de largo alcance, de los cuales a su vez 38 son contactos de enlace de hidrógeno. Para 1ENH contabilizamos 194 contactos nativos en total, de los cuales 99 son de largo alcance y 33 de estos son considerados contactos de enlace de hidrógeno.

En el mapa de contactos queda patente que para estas proteínas los enlaces de hidrógeno se dan exclusivamente a lo largo de las hélices, destacando la ausencia de enlaces de hidrógeno entre aminoácidos lejanos a lo largo de la secuencia.

4.2.2. Proteínas β

Para esta otra familia de proteínas existen dos tipos de estructuras principales. Como no es frecuente que las láminas β mantengan sus dos caras expuestas al disolvente, una primera posibilidad es que se agrupen en forma de sándwich, cuando dos láminas quedan unidas por sus caras interiores a través de aminoácidos hidrófobos, dejando la otra cara de cada lámina expuesta hacia al disolvente. La otra posibilidad es que las hebras β se agrupen en forma de barril, manteniendo un espacio en el interior, ocupado principalmente por cadenas laterales de naturaleza hidrófoba, y manteniendo

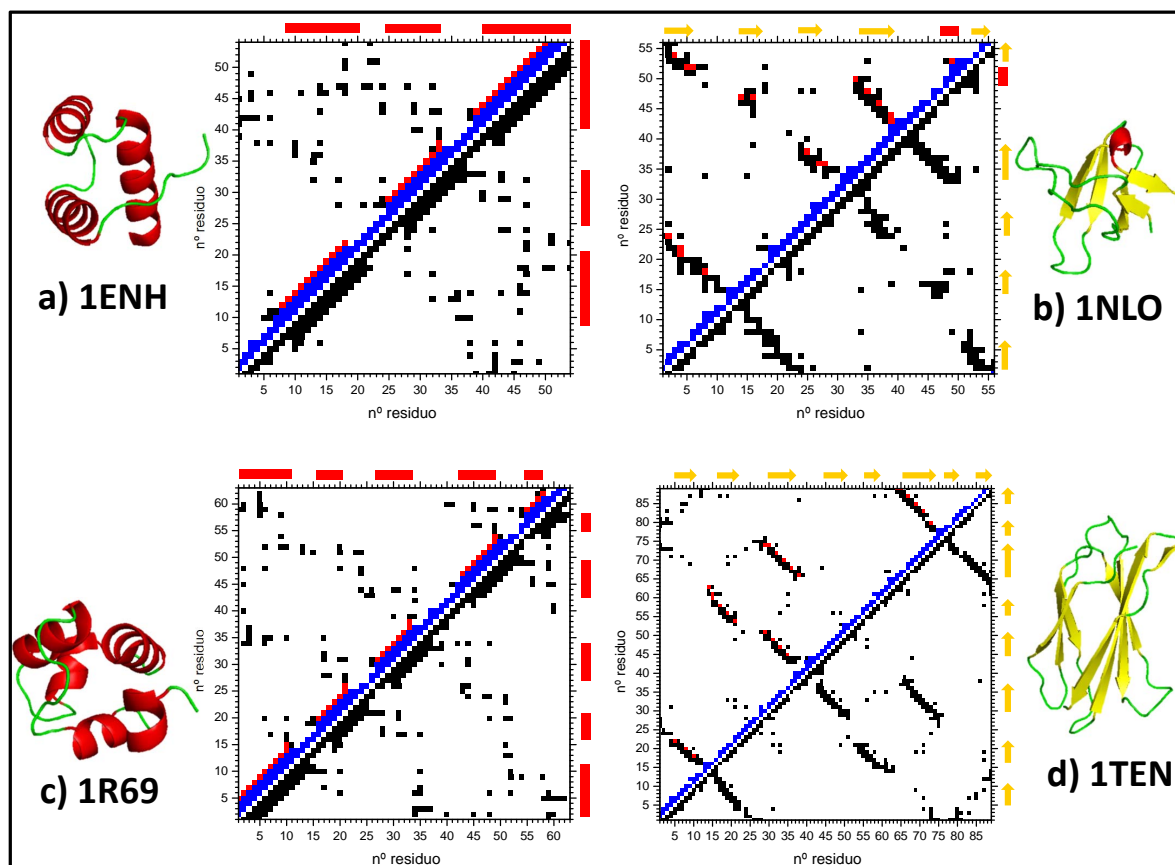


Figura 4.1: **Mapas de contactos para las cuatro proteínas.** Se muestra además una representación esquemática del estado nativo de cada una de las proteínas, para poder observar cómo se distribuyen los diferentes elementos de estructura secundaria. Los puntos azules del mapa corresponden a contactos locales, y los rojos a contactos de enlace de hidrógeno.

los aminoácidos hidrófilos hacia el disolvente. Este tipo de proteínas suele ser utilizado por los seres vivos para el transporte de sustratos hidrófobos a través de disolventes hidrófilos. Las proteínas introducen en el interior del sándwich o del barril el sustrato hidrófobo que se quiere transportar, y el exterior de la proteína queda tapizado por los aminoácidos polares, que hacen que la proteína sea soluble en el medio acuoso⁷.

De esta familia de proteínas hemos seleccionado el *Fibronectin type III domain*, 1TEN^{174,175} según su código PDB, que presenta un plegamiento típico de dos estados a presión atmosférica ampliamente estudiado^{176,177}. También estudiaremos la proteína *Src SH3 domain*, que corresponde a la estructura 1NLO¹⁷⁸ según su código PDB. Se trata de una proteína muy estudiada, debido a su relevancia biológica en el estudio de agregados¹⁷⁹, cuyo plegamiento se produce en un proceso de dos estados que sin

embargo resulta bastante rápido¹⁸⁰. Son dos proteínas que difieren en su tamaño, 89 aminoácidos para 1TEN y 56 aminoácidos en 1NLO, puesto que 1TEN está compuesta por 8 hebras β mientras que 1NLO está formada por 5 hebras β . La disposición espacial de dichas hebras se puede observar en los paneles (b) y (d) de la Figura 4.1, donde se presentan además los mapas de contactos de ambas proteínas. 1TEN presenta un total de 386 contactos, de los cuales 213 son de largo alcance y 25 son considerados contactos de enlace de hidrógeno. 1NLO presenta un total de 225 contactos, de los que 118 son de largo alcance y 18 son considerados contactos de enlace de hidrógeno. Pese a tratarse de proteínas de distinto tamaño, ambas presentan cinco horquillas β antiparalelas, a lo largo de las cuales aparecen los contactos de enlace de hidrógeno, dando lugar a mapas de contactos bastante similares en apariencia aunque de diferente tamaño.

4.2.3. Diferencias entre las dos familias de proteínas

Estas dos familias de proteínas presentan mapas de contactos con diferentes características. A continuación discutiremos las diferencias principales entre ellas, al mismo tiempo que describimos el comportamiento que podemos esperar para nuestro modelo en función de los mapas de contactos que presenta cada proteína. De nuevo queremos recordar que el mapa de contactos de cada proteína depende exclusivamente de su estructura nativa experimental. Por tanto, las diferencias que comentaremos a lo largo de esta Sección entre los dos principales elementos de estructura secundaria son una consecuencia del estado nativo y en ningún caso hemos introducido de forma artificial un comportamiento diferenciado entre los contactos que identifican a los dos motivos de estructura secundaria.

Cuando comparamos los cuatro mapas de contactos de la Figura 4.1, observamos dos diferencias principales entre las dos familias de proteínas. Por un lado presentan una proporción de contactos de enlace de hidrógeno bastante diferente, como ya hemos comentado anteriormente al mostrar las cifras en cada una de las estructuras. Pero además,

existe una distribución bien diferenciada del tipo de contactos que sirven para estabilizar los distintos elementos de estructura secundaria. Una estructura β requiere un mayor número de contactos de largo alcance para estabilizar la estructura secundaria, entre los cuales sólo algunos de ellos presentan enlace de hidrógeno. Las estructuras α , por otra parte, necesitan menos contactos, y de naturaleza muy local, para definir su estructura secundaria regular. Sin embargo, todos los residuos que participan en una hélice α se encuentran formando enlaces de hidrógeno con otros residuos de la misma hélice. Esta diferencia evidente en la distribución de los tipos de contactos, que se consideran de forma distinta en nuestra definición de las interacciones del modelo de interacción, debe ser el factor principal para que ambos tipos de estructuras presenten un tránsito de plegamiento/desnaturalización diferenciado a medida que aumenta la presión.

Las estructuras α deberían de ser capaces de estabilizar, al menos parcialmente, sus hélices por separado, y mantener estas unidas en la estructura nativa a través de contactos de largo alcance, susceptibles de verse afectados por un aumento de presión. Por lo tanto, en las hélices α hay una diferencia clara entre los contactos que estabilizan la estructura secundaria y los que estabilizan la estructura terciaria de la proteína. En las estructuras β , por el contrario, debe producirse un acercamiento espacial de aminoácidos lejanos en la secuencia para estabilizar la estructura secundaria, sobre todo cuando se forman contactos entre hebras β muy alejadas a lo largo de la secuencia. Por lo tanto, la formación de la estructura secundaria obliga al empaquetamiento de toda la proteína. En estas estructuras, por consiguiente, se hace mucho más difícil la distinción entre los contactos que estabilizan la estructura secundaria y los que estabilizan la estructura terciaria.

De acuerdo a las características de nuestro potencial de interacción dependiente de la presión, las hélices α están únicamente formadas por contactos de corto alcance y contactos de enlace de hidrógeno. Por tanto, una hélice α individual en ningún caso podrá intercalar entre los aminoácidos que la constituyen una molécula de agua, dando

lugar a una ‘hélice α hinchada’. El tipo de contactos presentes hace que esta situación no se estabilice en nuestro modelo. De esta manera, la estructura de la hélice α puede estar formada o no formada, y lo que realmente puede ‘hincharse’ con nuestro modelo al aumentar la presión es la interacción entre diferentes hélices, debida a los los contactos que estabilizan la estructura terciaria. Esto nos llevaría a una estructura ‘hinchada’ donde realmente las hélices pueden estar perfectamente formadas, en una conformación local muy parecida a la del estado nativo, mientras que la distancia entre ellas puede aumentar en aproximadamente 3 Å (la diferencia entre los mínimos de contacto directo y contacto hinchado en nuestro potencial). Este tipo de estructura hinchada debería generar poca frustración al sistema, puesto que a priori parece ser una estructura a la que nuestra cadena polipeptídica podría adaptarse con los diferentes contactos en mínimos del potencial.

En las proteínas β (Figura 4.1) observamos cómo los contactos nativos que sirven para estabilizar los elementos de estructura secundaria se pueden encontrar como contactos de enlace de hidrógeno o como contactos de largo alcance, totalmente mezclados entre sí. Esta distribución heterogénea de contactos hace que el sistema sufra una mayor frustración al aplicarle nuestro potencial de interacción. Consideremos por ejemplo un contacto nativo (i, j) que corresponde a un enlace de hidrógeno, mientras que los contactos $(i - 1, j + 1)$ o $(i + 1, j - 1)$ son contactos de largo alcance. Al aumentar la presión, estos últimos intentarán pasar al mínimo hinchado, mientras que el primero, para el que este mínimo no existe, intentará permanecer en el mínimo de contacto directo. El hecho de que estemos ante una cadena polipeptídica, donde la secuencia covalente no va a romperse por un incremento de presión, genera mucha frustración al sistema cuando ϵ_{ssm} adquiere valores iguales o mayores que la unidad. En principio, dado el mayor número de contactos clasificados como de largo alcance sin enlace de hidrógeno, al aumentar la presión se favorece una estructura donde los diferentes residuos pueden formar una estructura β ‘hinchada’, donde la distancia entre hebras haya aumentado en

3 Å aproximadamente. Esta situación es poco realista desde un punto de vista físico (los enlaces de hidrógeno entre hebras β desaparecerían, lo que pondría en serio compromiso la estabilidad conformacional de estas mismas hebras), pero puede aparecer en nuestro modelo. Como alternativa, en un intento de encontrar una situación de equilibrio entre los contactos directos de enlace de hidrógeno y los contactos de largo alcance hinchados el modelo nos podría conducir al aumentar la presión a una deformación variable de la estructura β en el estado hinchado, puesto que la cadena podría no encontrar una única conformación estable.

A lo largo de la próxima Sección comprobaremos si las diferencias que presentan los mapas de contactos son suficientes para que nuestro modelo trate de diferente manera a las proteínas α y a las proteínas β .

4.3. Resultados al aplicar nuestro potencial de interacción

A continuación mostramos los resultados más relevantes que hemos obtenido al realizar la simulación de las cuatro proteínas mencionadas con nuestro modelo. Como sucedía en el Capítulo anterior, la metodología utilizada para estudiar el efecto del aumento de la presión hidrostática en el medio se basa en realizar diferentes simulaciones haciendo variar el parámetro ϵ_{ssm} . El valor máximo que adopta este parámetro para cada proteína depende del sistema concreto, puesto que realizamos simulaciones a valores crecientes de dicho parámetro hasta que observemos que nos encontramos en un régimen de altas presiones donde incluso a la menor temperatura de la simulación no detectamos la presencia del estado nativo. En la Tabla 4.1 se muestra un resumen de algunas de las propiedades de las cuatro proteínas, junto con los intervalos de valores de ϵ_{ssm} a los que ha sido simulada cada una de ellas. Precisamente la gran cantidad de simulaciones realizadas para obtener los resultados de este Capítulo es la que nos obliga a presentar

solo aquellos que consideramos más importantes y representativos para cada sistema.

	1ENH	1R69	1NLO	1TEN
Tipo de proteína	α	α	β	β
nº de aminoácidos	54	69	56	89
Intervalo de ϵ_{ssm} simulado	0.00 - 1.40	0.00 - 1.40	0.00 - 1.50	0.00 - 1.40

Tabla 4.1: Resumen de las proteínas simuladas.

4.3.1. Tipos de plegamiento observado

Como se mostró en el Capítulo 3, el primer paso para caracterizar el proceso de desnaturalización que presenta una proteína es la representación de sus curvas de capacidad calorífica frente a la temperatura, y de sus perfiles de energía de Helmholtz a la temperatura de transición. En la Figura 4.2 se presentan las curvas de capacidad calorífica que hemos obtenido para los cuatro sistemas estudiados a los diferentes valores de ϵ_{ssm} a los que hemos realizado las simulaciones.

Podemos observar cómo la evolución de las curvas con el aumento del parámetro ϵ_{ssm} es similar para todas las proteínas, y es un comportamiento muy parecido al que habíamos obtenido para la proteína 2GB1. Cuando la presión del sistema aumenta, existe un claro desplazamiento de la curva hacia menores temperaturas de transición (indicadas por el máximo de cada una de ellas). Este comportamiento se invierte para los valores más altos de ϵ_{ssm} , condiciones de alta presión, del mismo modo que sucedía con la proteína 2GB1. La proteína 1TEN es la única que en su régimen de bajas presiones presenta un comportamiento diferente al indicado, al aumentar ligeramente su temperatura de transición a medida que aumenta el valor de ϵ_{ssm} . Posiblemente el mayor tamaño de esta proteína, la más grande con diferencia de todas las consideradas, incluyendo 2GB1, está relacionada con este comportamiento a presiones bajas.

Además del desplazamiento de la temperatura de transición, con el aumento de la presión observamos también una evolución en el tipo de tránsito desde perfiles de

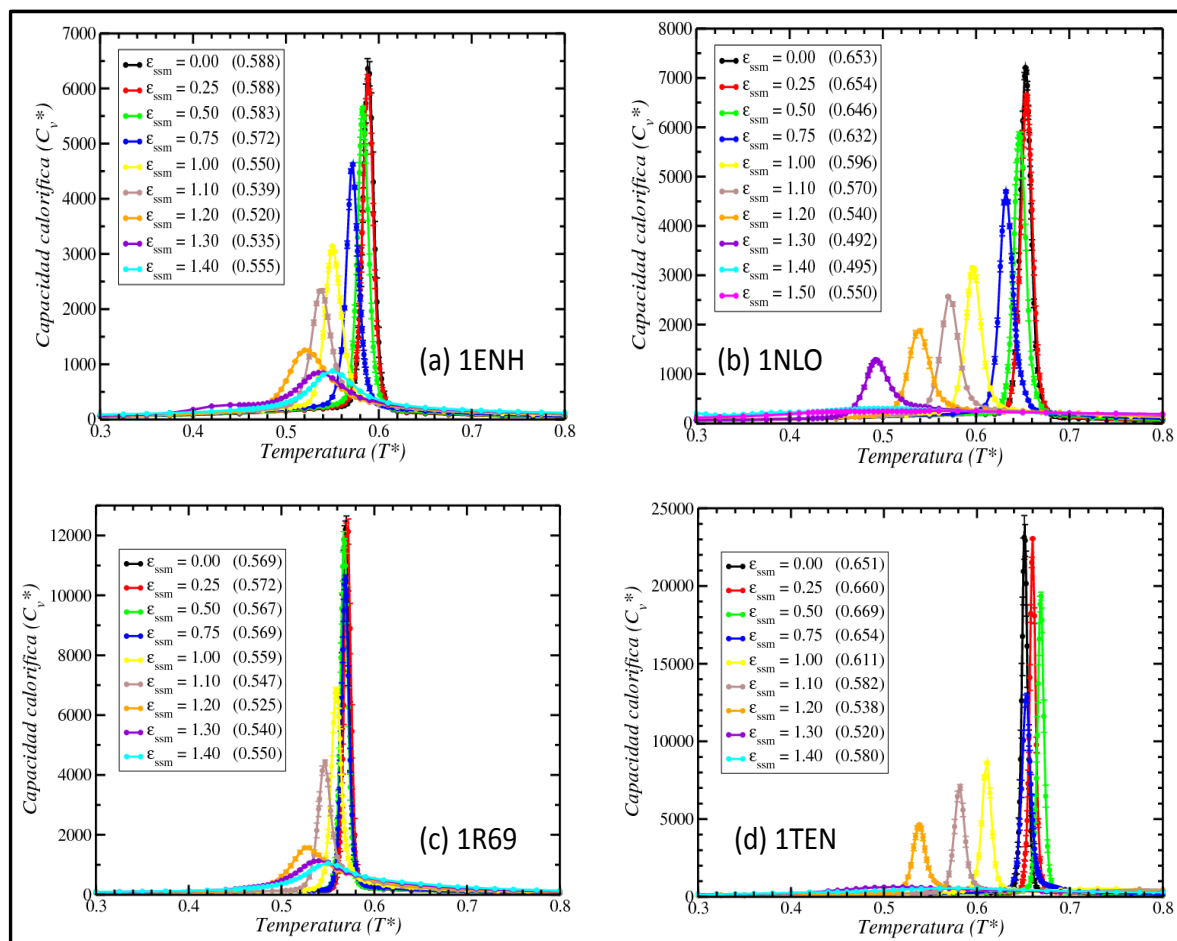


Figura 4.2: **Curvas de capacidad calorífica de todas las simulaciones.** a) Proteína α 1ENH. b) Proteína β 1NLO. c) Proteína α 1R69. d) Proteína β 1TEN. En la leyenda se muestran los diferentes valores de ϵ_{ssm} a los que ha sido realizada cada simulación. El valor entre paréntesis indica la temperatura reducida correspondiente al máximo de cada curva.

capacidad calorífica estrechos y altos, correspondientes a tránsitos cooperativos, hasta tránsitos muy poco cooperativos, correspondientes a perfiles anchos y bajos. Además, si analizamos las representaciones de la capacidad calorífica en función del tipo de estructura que presenta cada proteína, podemos extraer algunas conclusiones relacionadas con el tipo de estructura secundaria. Los paneles de la izquierda de la Figura 4.2, (a) y (c), corresponden a las proteínas α mientras que los de la derecha, (b) y (d), corresponden a las proteínas β . Las dos diferencias fundamentales que se observan entre ambas familias estructurales son:

1. **El desplazamiento de la temperatura de transición, a medida que aumenta la presión, es mayor en las proteínas β que en las α .** Se observa claramente cómo para las dos proteínas β existe un desplazamiento en la temperatura de transición mucho más pronunciado del que observamos para las proteínas α a medida que aumenta el valor del parámetro ϵ_{ssm} . De hecho, si medimos el desplazamiento relativo de la temperatura de transición desde que $\epsilon_{ssm} = 0.00$ hasta que $\epsilon_{ssm} = 1.20$ para las cuatro proteínas obtenemos que las proteínas α , 1ENH y 1R69, reducen en un 11 % y un 8 % su valor de T_m , respectivamente, mientras que en las proteínas β , 1NLO y 1TEN, los desplazamientos son de aproximadamente un 17 % en ambos casos. Este mayor desplazamiento en la temperatura de desnaturalización de las proteínas β nos indica que este tipo de estructura secundaria se ve más afectado por el efecto de la presión que las estructuras α , lo que parece implicar que con nuestro modelo a medida que aumenta la presión del medio las hélices α se desestabilizan en menor medida que las láminas β . Este resultado es particularmente interesante, ya que coincide con las observaciones experimentales mencionadas más arriba^{93,166}.

2. **Existen diferencias significativas en las curvas correspondientes al régimen de altas presiones.** Si bien las características específicas que identifican a cada régimen de presiones deben ser identificadas para cada uno de los nuevos sistemas a lo largo de esta Sección, observamos cómo para los valores más altos del parámetro ϵ_{ssm} existe un comportamiento diferenciado dependiendo del tipo de proteína que estudiemos. Las proteínas α presentan curvas muy poco cooperativas a estas presiones, pero al menos presentan un perfil típico de pico ensanchado. Sin embargo, las proteínas β presentan un tránsito a altas presiones donde apenas aparece ningún tipo de pico, y de hecho no está claro que en estas condiciones se esté detectando un tránsito en las curvas de capacidad calorífica para estas proteínas β . De nuevo, esto parece indicar que el efecto que ejerce la presión es más

acusado en las proteínas β que en las α . Además, en ambos casos se observa que existe un valor límite de ϵ_{ssm} por encima del cual el comportamiento de la curva presenta un cambio drástico hacia curvas típicas de tránsitos *downhill*, siendo más acusado en las proteínas β que en las proteínas α .

Los perfiles de energía de Helmholtz a las temperaturas de los correspondientes máximos de capacidad calorífica se muestran en la Figura 4.3. En ellos se observa cómo para las cuatro proteínas estudiadas en este Capítulo existe una evolución del tipo de tránsito muy similar al que presenta la proteína 2GB1. A medida que aumenta la presión del sistema, la barrera de energía de Helmholtz existente entre el estado nativo y el estado desnaturalizado va disminuyendo, al mismo tiempo que el estado desnaturalizado va evolucionando hacia una situación de menor energía, dando lugar a presiones moderadas y altas a la aparición del nuevo estado desnaturalizado. Además, también observamos cómo por encima de un determinado valor de ϵ_{ssm} , alrededor de 1.20 para las cuatro proteínas, existe un cambio de tendencia y se pasa de presentar dos estados en equilibrio a la temperatura de transición, separados por una barrera más o menos grande, a estabilizar un único estado, de energía intermedia entre ambos.

Si bien los perfiles presentan tendencias similares para todas las proteínas, encontramos una característica que diferencia a las dos familias estructurales. Observamos cómo, para los valores más altos de ϵ_{ssm} , todas las proteínas presentan el perfil típico de un tránsito *downhill*. Sin embargo las proteínas α , cuyos resultados aparecen en los paneles (a) y (c), presentan en estas condiciones de alta presión un perfil mucho más ancho que además prácticamente se estabiliza en un valor de energía que permanece constante aunque ϵ_{ssm} siga aumentando, mientras que las proteínas β presentan en estas presiones un perfil más estrecho, cuya energía continúa evolucionando con ϵ_{ssm} , paneles (b) y (d). Esto parece indicar que a altas presiones el estado desnaturalizado de las proteínas β se mantiene todavía afectado por la presión, mientras que las proteína α alcanzan un estado estable a la temperatura de transición que no se ve modificado por aumentos

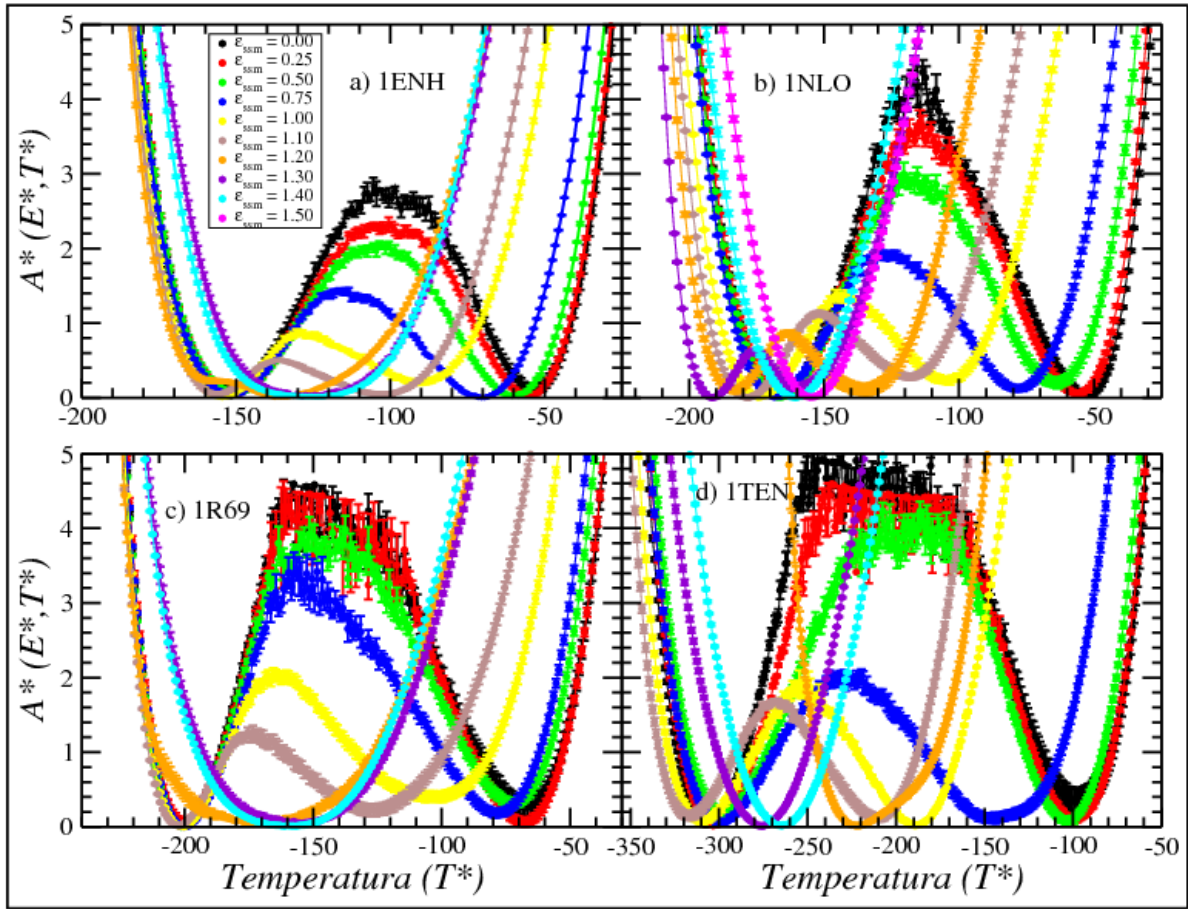


Figura 4.3: **Perfiles de energía de Helmholtz a la temperatura de transición.** Se presentan los diferentes perfiles de energía de Helmholtz, cada uno calculado a la temperatura del máximo de la correspondiente curva de capacidad calorífica de la Figura 4.2.

posteriores de la presión.

Por tanto, gracias a las curvas de capacidad calorífica y a los perfiles de energía de Helmholtz, podemos afirmar que las cuatro proteínas estudiadas presentan transiciones que globalmente se pueden considerar similares a la que habíamos presentado para la proteína 2GB1, con la presencia de dos estados a la temperatura de transición cuando nos encontramos a presión atmosférica. A presiones moderadas aparece un tránsito de desnaturalización poco cooperativo entre un estado nativo y un estado desnaturalizado que difiere del estado desnaturalizado clásico a presión baja. Y cuando nos encontramos a altas presiones encontramos un tránsito continuo entre dos estados desnaturalizados diferentes.

4.3.2. Evolución de los diferentes estados

Al igual que describimos en la Capítulo 3, nos ayudaremos de los histogramas mono y bidimensionales de las propiedades estructurales registradas a lo largo de las simulaciones para así analizar las diferentes conformaciones registradas. Así, tras un análisis de los tres estados a diferentes temperaturas y presiones, seleccionamos los intervalos de radio de giro reducido y RMSD característicos de cada estado. En la Tabla 4.2 mostramos los intervalos que caracterizan a cada uno de los tres estados para las cuatro proteínas estudiadas. Los valores característicos para cada estado son, como cabía esperar, distintos en cada uno de los sistemas. Esto se debe a que, aunque cada una de las proteínas estabiliza estados con características similares (nativo, hinchado y desnaturalizado), cada uno de estos estados es diferente en función del tamaño y las características estructurales del estado nativo de cada proteína, y por tanto debemos seleccionar un intervalo específico para cada una de ellas. Una vez seleccionado el intervalo de radio de giro reducido y RMSD que caracteriza cada estado, podemos realizar un estudio cuantitativo de sus poblaciones para cada una de las cuatro proteínas, que nos permitirá estudiar de nuevo los tres tipos de transición encontrados a lo largo de la simulación: N-D, N-H e H-D. Además, gracias a las gráficas de poblaciones, que mostramos en las Figuras 4.4 y 4.5, podemos definir los valores de ϵ_{ssm} que identifican aproximadamente los diferentes regímenes de presión para cada una de las cuatro proteínas. Los intervalos

	Estado nativo	Estado hinchado	Estado desnaturalizado
1ENH	$R_g^* < 1.035$ $RMSD < 3.0 \text{ \AA}$	$1.035 < R_g^* < 1.05$ $3.0 < RMSD < 7.0 \text{ \AA}$	$R_g^* > 1.05$ $RMSD > 7.0 \text{ \AA}$
1R69	$R_g^* < 1.020$ $RMSD < 2.5 \text{ \AA}$	$1.020 < R_g^* < 1.10$ $2.5 < RMSD < 7.0 \text{ \AA}$	$R_g^* > 1.10$ $RMSD > 7.0 \text{ \AA}$
1NLO	$R_g^* < 1.040$ $RMSD < 3.5 \text{ \AA}$	$1.040 < R_g^* < 1.20$ $3.5 < RMSD < 8.0 \text{ \AA}$	$R_g^* > 1.20$ $RMSD > 8.0 \text{ \AA}$
1TEN	$R_g^* < 1.020$ $RMSD < 2.0 \text{ \AA}$	$1.020 < R_g^* < 1.10$ $2.0 < RMSD < 10.0 \text{ \AA}$	$R_g^* > 1.10$ $RMSD > 10.0 \text{ \AA}$

Tabla 4.2: Valores límite de radio de giro (reducido con el valor de la estructura nativa) y RMSD que identifican a cada estado para las cuatro proteínas.

de ϵ_{ssm} característicos de cada estado, para las cuatro proteínas estudiadas, se recogen en la Tabla 4.3. Podemos observar en esta Tabla que los intervalos que caracterizan a cada régimen de presión son muy similares para todas las proteínas, lo que favorece la existencia de una relación entre el valor de ϵ_{ssm} y la presión hidrostática, como estamos considerando a lo largo de esta Tesis. Para las cuatro proteínas, las curvas de poblaciones en función de la temperatura nos muestran, como ocurría con 2GB1, la presencia de un equilibrio de dos estados entre el estado nativo y el estado desnaturalizado cuando nos encontramos en un régimen de bajas presiones. A presiones moderadas observamos la presencia de un tránsito menos cooperativo que el anterior entre el estado nativo y el estado hinchado, y de un tránsito muy poco o nada cooperativo entre el estado hinchado y el estado desnaturalizado. Finalmente, a altas presiones sólo observemos un tránsito

	Bajas presiones	Presiones moderadas	Altas presiones
1ENH	$\epsilon_{ssm} \leq 0.75$	$\epsilon_{ssm}: 1.00 - 1.20$	$\epsilon_{ssm} \geq 1.30$
1R69	$\epsilon_{ssm} \leq 1.00$	$\epsilon_{ssm}: 1.10 - 1.20$	$\epsilon_{ssm} \geq 1.30$
1NLO	$\epsilon_{ssm} \leq 0.75$	$\epsilon_{ssm}: 1.00 - 1.30$	$\epsilon_{ssm} \geq 1.40$
1TEN	$\epsilon_{ssm} \leq 0.75$	$\epsilon_{ssm}: 1.00 - 1.20$	$\epsilon_{ssm} \geq 1.30$

Tabla 4.3: Regímenes de presión en función de ϵ_{ssm} para las cuatro proteínas.

no cooperativo entre el estado hinchado y el estado desnaturalizado. De nuevo, las características mencionadas para los diferentes tránsitos se basan en las propiedades de las curvas representadas en las regiones en las que se produce el cruce entre poblaciones de dos estados diferentes. Así, observamos que las pendientes de las curvas que representan a cada estado van disminuyendo a medida que aumenta el valor de ϵ_{ssm} , lo cual indica que al aumentar la presión todos los tránsitos reducen notablemente su cooperatividad. Esto corrobora de nuevo lo que habíamos apreciado en los perfiles de energía de Helmholtz, al disminuir la barrera energética que separa los diferentes estados que aparecen a lo largo del proceso de desnaturalización cuando ϵ_{ssm} aumenta.

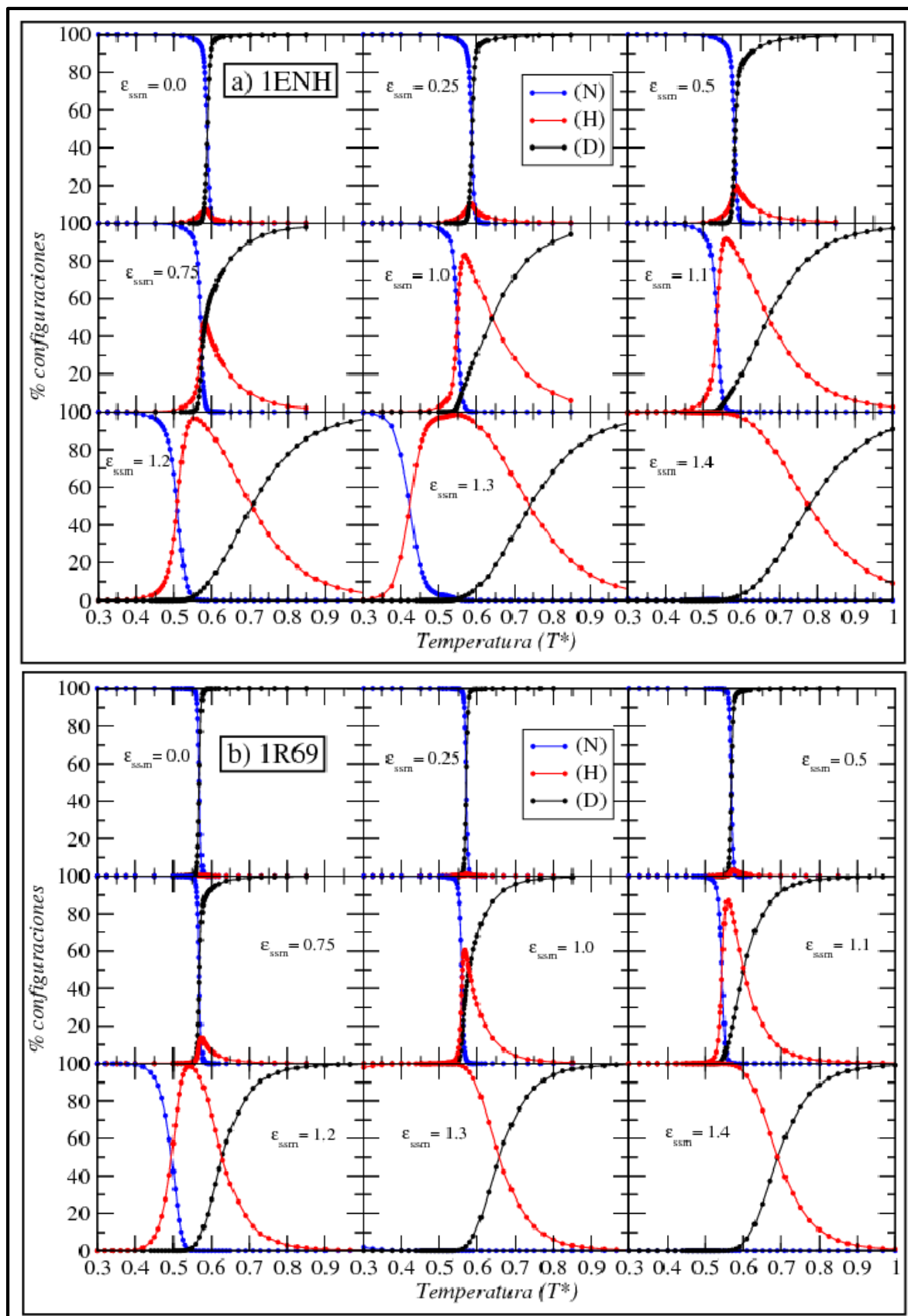


Figura 4.4: Diagramas de poblaciones de las proteínas α . a) Proteína 1ENH. b) Proteína 1R69.

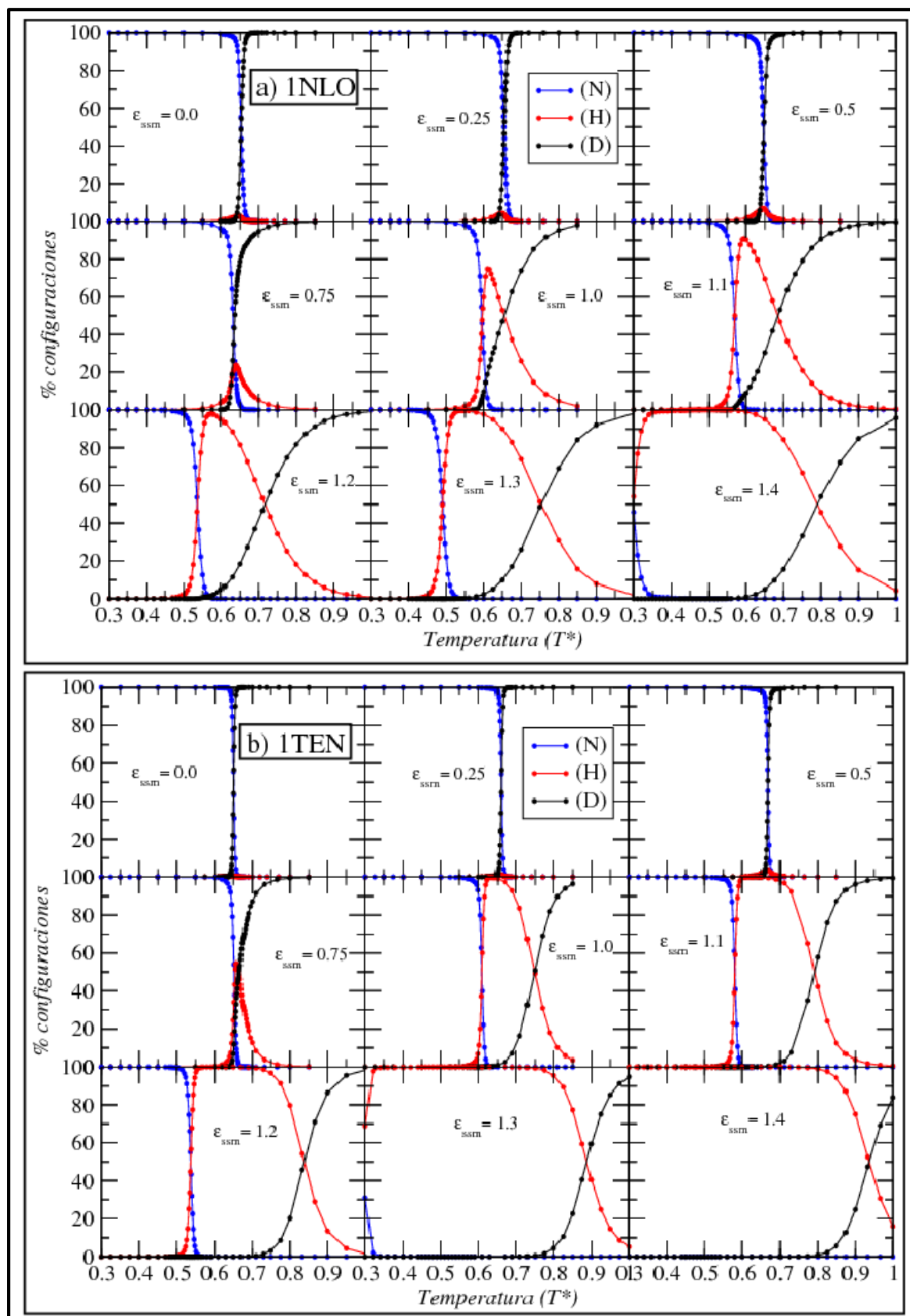


Figura 4.5: Diagramas de poblaciones de las proteínas β . a) Proteína 1NLO. b) Proteína 1TEN.

4.3.3. Análisis de los estados hinchados

Una vez analizados, al menos de forma cualitativa, los diferentes tránsitos que se dan en cada sistema, decidimos realizar un estudio estructural del estado hinchado que aparece en cada una de las proteínas. De esta manera comprobaremos sus características de acuerdo a nuestro modelo de simulación. Para el análisis de los diferentes estados hinchados utilizaremos los mapas de frecuencia de contactos que presentamos en el Capítulo anterior, para así poder estudiar los aspectos estructurales más relevantes que caracterizan las diferentes conformaciones registradas. En la Figura 4.6 mostramos cuatro mapas de frecuencias de contactos, que representan el estado hinchado de cada una de las cuatro proteínas consideradas en este Capítulo. Cada uno de los mapas de frecuencia de contactos ha sido obtenido a una temperatura diferente, ya que cada mapa ha sido calculado para los valores menores de ϵ_{ssm} y temperatura en los que el estado hinchado se encuentra poblado en al menos un 75 % de los casos registrados en la simulación correspondiente. En estas condiciones el estado hinchado es suficientemente estable, a tenor de los diagramas de poblaciones que hemos obtenido. Sin embargo, nos encontraremos en unas condiciones en las que la presión todavía es moderada. Además, al seleccionar la temperatura más baja para la que el estado hinchado es estable, reducimos todo lo posible las fluctuaciones térmicas del sistema, lo que aporta una información más consistente a los valores promedio que se representan en los mapas de frecuencias de contactos. De esta forma, en la Figura podemos apreciar los primeros estados hinchados que obtenemos cuando la presión empieza a ejercer un efecto de desnaturalización. Por último, como en este análisis el total de las conformaciones obtenidas en las condiciones de simulación no corresponden al estado hinchado, hemos tenido que seleccionar en función de los valores de radio de giro reducido y RMSD, que mostramos en la Tabla 4.2, las conformaciones que pertenecen al mismo, y calcular las frecuencias de contactos sólo sobre ellas.

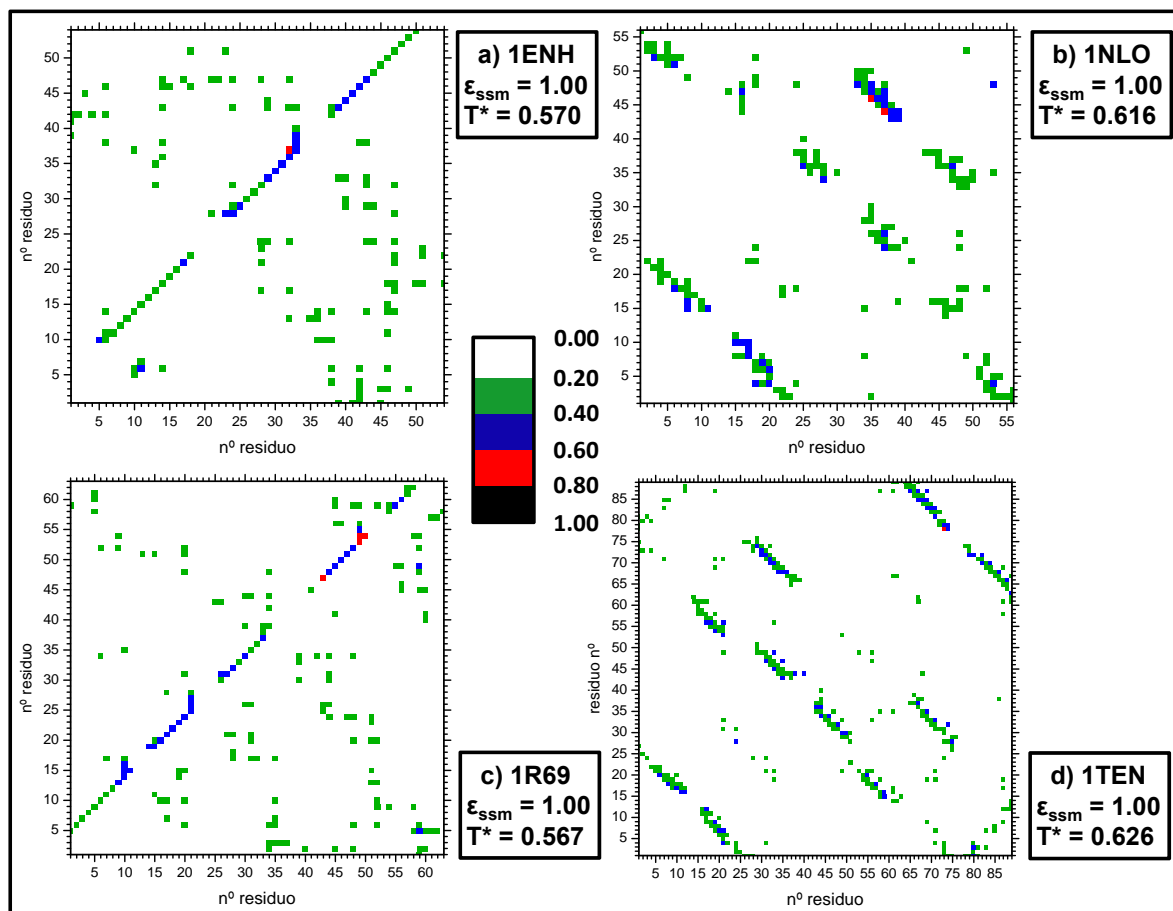


Figura 4.6: **Mapas de frecuencia de contactos de los estados hinchados a presiones moderadas.** Mapas de frecuencia de contactos para las cuatro proteínas estudiadas en condiciones de presión moderada. Cada mapa está realizado a una temperatura diferente, señalada en la correspondiente leyenda. Como ya explicamos en la Sección 3.8, el triángulo superior izquierdo muestra la frecuencia de contactos directos, mientras que el triángulo inferior derecho muestra la frecuencia de contactos hinchados.

En la Figura apreciamos la presencia en las condiciones mostradas de conformaciones con una estructura nativa apenas definida, ya que la mayoría de los contactos nativos muestran frecuencias promedio de aparición muy pequeñas. Entre los contactos nativos con frecuencia significativa, observamos claras diferencias entre las proteínas α , paneles a) y c), y las β , paneles b) y d). Para las primeras, los contactos de mayor población corresponden a contactos directos que representan los enlaces de hidrógeno en las hélices α de estas proteínas. Como comentábamos al principio de este Capítulo, esta es la situación esperada con nuestro modelo, ya que estos contactos no estabilizan contactos hinchados. Las diferentes hélices presentes muestran estabilidades diferentes, aunque

los resultados concretos dependen mucho de las condiciones concretas escogidas para el análisis. En las proteínas β , por el contrario, aproximadamente la mitad de los contactos con probabilidades significativas se encuentren como contactos directos y la otra mitad como contactos hinchados. Son contactos que aparecen en las zonas de interacción entre diferentes hebras β asociadas de forma antiparalela. Aquí observamos claramente el efecto de la frustración que comentábamos en la introducción de este Capítulo. Si bien los enlaces de hidrógeno existentes intentan mantener la estructura secundaria en la región de los contactos directos, el mayor número de contactos nativos de largo alcance favorece la posibilidad de estabilizar contactos hinchados, incluso a costa de perder alguno de los enlaces de hidrógeno. Las conformaciones registradas alternan entre estas dos situaciones, por lo que en promedio se obtienen resultados como los recogidos en la Figura 4.6.

Globalmente, estas distribuciones de los contactos nativos nos indican que en las condiciones en las que se ha obtenido la Figura 4.6 existe un aumento de la presión suficientemente alto para que la conformación nativa de la proteína se vea sustancialmente alterada, pero sin que se establezca un estado hinchado con características estructurales suficientemente bien definidas. La única conclusión relevante, que confirma lo que ya habíamos visto al estudiar las propiedades termodinámicas de estas proteínas, es que la presión desestabiliza en mayor medida a la estructura nativa β que a la α .

Para seguir entendiendo cómo evolucionan los estados hinchados a medida que aumenta la presión del sistema, mostramos en la Figura 4.7 los mapas de frecuencias de contactos pertenecientes al estado hinchado que hemos obtenido dentro de un régimen de altas presiones, $\epsilon_{ssm} = 1.30$, y a una temperatura moderada. En este caso, además, hemos preferido mantener la misma temperatura para todas las conformaciones, $T^* = 0.550$, de tal forma que las fluctuaciones térmicas sean similares para las cuatro proteínas en estas condiciones. En este caso, encontramos que la diferencia mencionada anteriormente entre las proteínas con estructura α y las β es mucho más acusada.

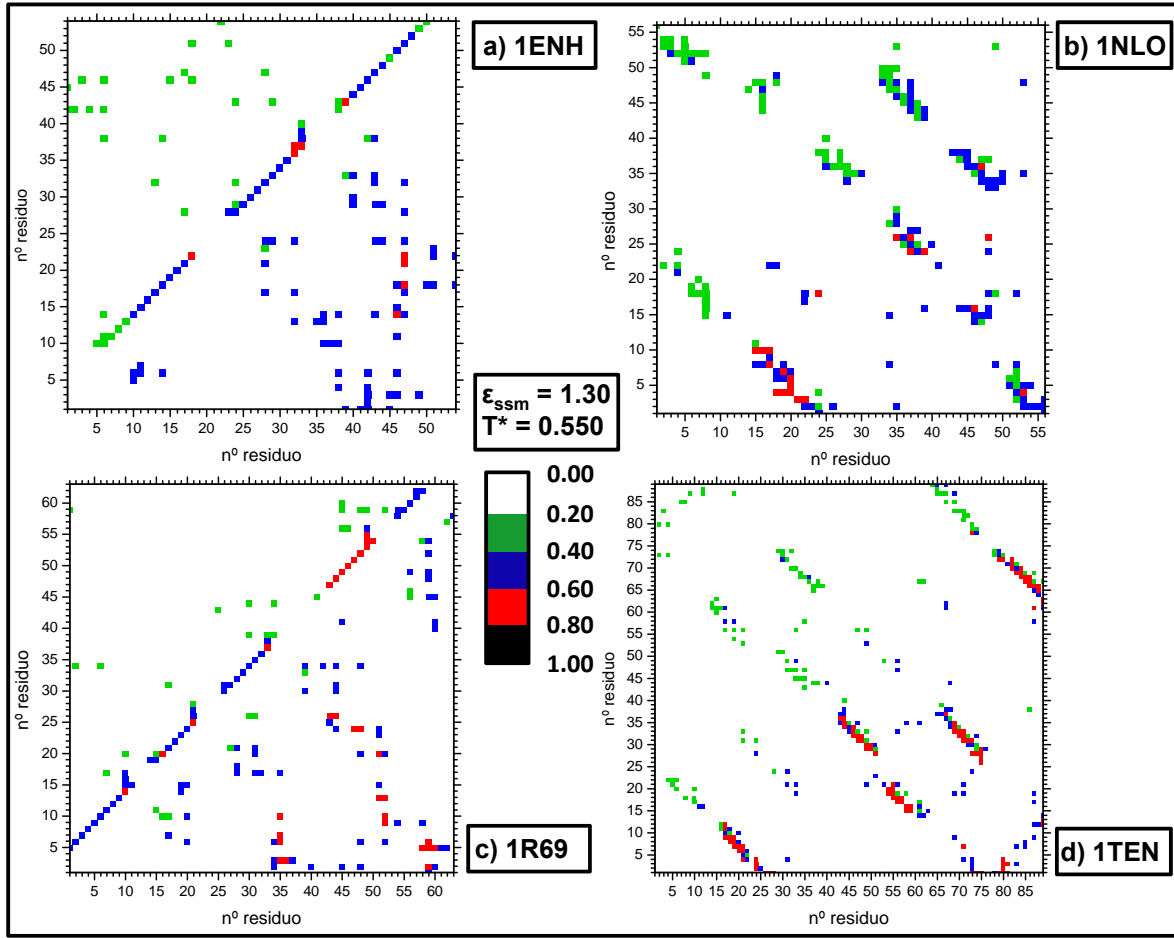


Figura 4.7: **Mapas de frecuencia de contactos de los estados hinchados a altas presiones.** Mapas de frecuencia de contactos para los cuatro estados hinchados dentro de un régimen de altas presiones, $\epsilon_{ssm} = 1,30$. Todos los mapas se presentan a la misma temperatura, $T^* = 0,550$.

Para las dos proteínas α , paneles (a) y (c), observamos un mapa donde los contactos que estabilizan las hélices se encuentran mayoritariamente como contactos directos y los contactos entre hélices que estabilizan la estructura terciaria globular se encuentran mayoritariamente como contactos hinchados. Recordamos que, como ya hemos venido comentando a lo largo de esta memoria, esta distribución asimétrica de los contactos nativos en función de si estabilizan la estructura secundaria o la terciaria de las proteínas es característico del estado hinchado desnaturalizado por presión^{39,49,50}. Nuestro modelo, por lo tanto, reproduce este estado de forma notable. Sin embargo, para las proteínas β , paneles (b) y (d), observamos una distribución diferente. Como ya hemos mencionado, los contactos entre hebras β alternan enlaces de hidrógeno, que estabilizan

los elementos de estructura secundaria, con contactos de largo alcance que estabilizan la estructura terciaria del barril β . La naturaleza covalente de la cadena polipeptídica impide en este caso que cada uno de los contactos se mantenga en su mínimo más estable. Por eso, muchos de los contactos nativos de largo alcance aparecen con población mayoritaria en el estado hinchado, que para este valor de ϵ_{ssm} es ya considerablemente más estable que el del contacto directo, rompiendo los contactos de enlace de hidrógeno correspondientes. Desde este punto de vista, es discutible afirmar si el estado desnaturalizado obtenido con nuestro modelo para proteínas β preserva o no la estructura secundaria nativa. Las hebras β están formadas, e interaccionan entre ellas formando el equivalente a una lámina β abierta, con separación excesiva entre sus componentes (unos 3 Å más que lo que presentan en el estado nativo), y que muy posiblemente no sería estable en una proteína real. De todas formas, dentro de la resolución reducida de nuestro modelo, esta situación de lámina β distorsionada es la que corresponde al estado desnaturalizado por presión, y todavía preserva la idea general de una estructura en la que los elementos de estructura secundaria están presentes, pero peor empaquetados como consecuencia de la inyección de moléculas de agua que da lugar a un aumento en el radio de giro de la estructura, que sigue manteniendo características esencialmente globulares. Por lo tanto, el estado desnaturalizado por presión en estas estructuras sigue siendo muy diferente del estado conformacionalmente desordenado correspondiente a la desnaturalización térmica.

Para poder estudiar la evolución de las estructuras β con la presión, ya que al menos en nuestro modelo parece ser más interesante que en las proteínas α , mostramos también el análisis de las conformaciones encontradas en las simulaciones de nuestras dos proteínas β a valores intermedios de ϵ_{ssm} . En la Figura 4.8 mostramos los mapas de frecuencia de contactos obtenidos cuando $\epsilon_{ssm} = 1.10$, paneles a) y b), y 1.20, paneles c) y d), con $T^* = 0.590$ en el caso de la proteína 1NLO, paneles (a) y (c), y $T^* = 0.610$ en el caso de la proteína 1TEN, paneles (b) y (d).

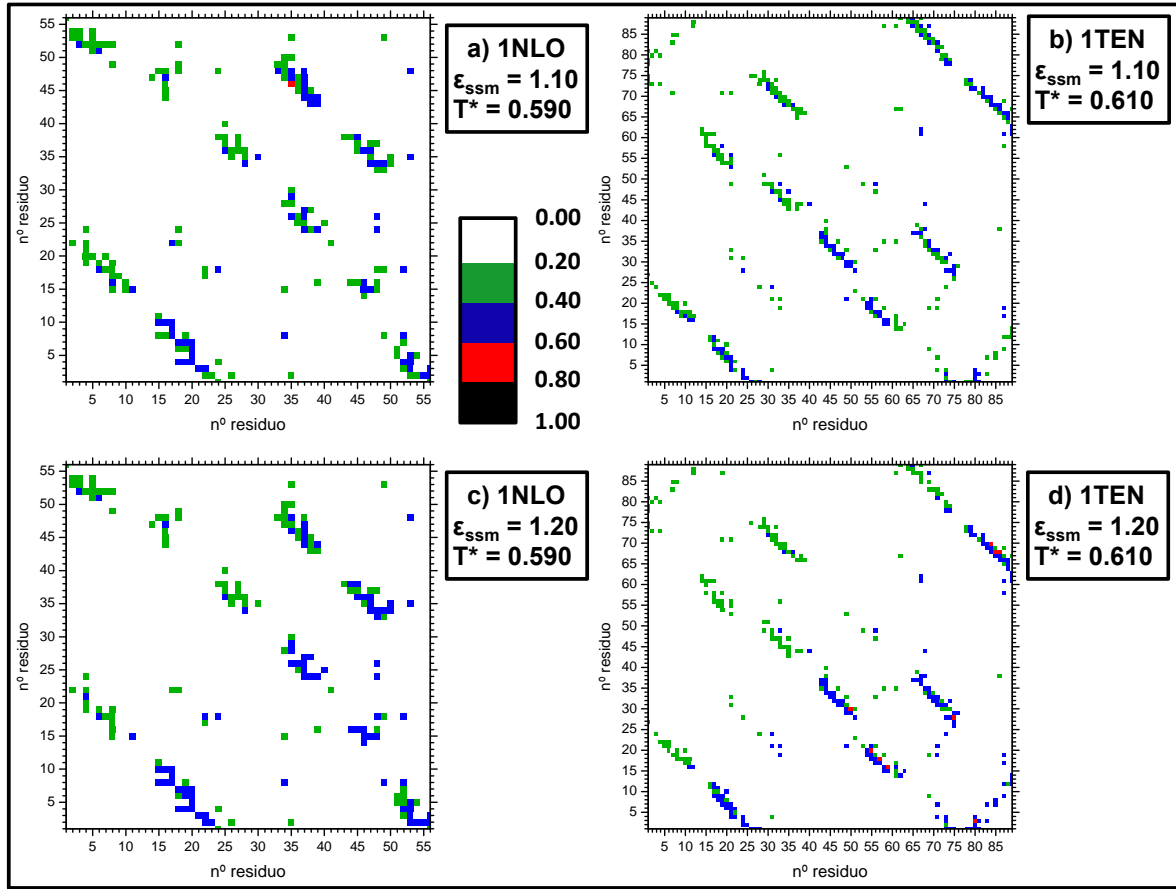


Figura 4.8: **Evolución del estado hinchado para proteínas β .** Se muestra la evolución de los mapas de frecuencia de contactos de las dos proteínas β a medida que crece el valor ϵ_{ssm} .

Añadiendo los mapas para $\epsilon_{ssm} = 1.0$ de la Figura 4.6 y los de $\epsilon_{ssm} = 1.3$ de la Figura 4.7, podemos observar con claridad cómo los sistemas con estructura β van resolviendo paulatinamente la frustración energética debida a la asimetría en las interacciones mencionada anteriormente. Para los valores de presión mostrados en la Figura 4.8, se observa con claridad el paso desde situaciones donde los contactos nativos entre hebras vecinas de la estructura fluctúan entre el mínimo de contacto directo y el de contacto hinchado, debido al efecto de los enlaces de hidrógeno, a un claro predominio de los contactos hinchados. Este paso no se produce de forma homogénea para todas las horquillas, según sea la proporción de contactos de largo alcance y de enlace de hidrógeno en cada una de ellas. Otros aspectos entrópicos difíciles de cuantificar, relacionados con la separación entre los residuos interaccionantes a lo largo de la secuencia, pueden jugar

también un papel en esta transición.

La necesidad de ir superando de forma progresiva las restricciones al hinchamiento impuestas por los enlaces de hidrógeno justifica la evolución continua de las propiedades termodinámicas con la presión que discutimos para las proteínas β al analizar las curvas de capacidad calorífica y los perfiles de energía de Helmholtz. En las estructuras α , donde los contactos de largo alcance que pasan al estado hinchado no compiten con los enlaces de hidrógeno que contribuyen a la estabilidad de las hélices, esta situación se resuelve de forma mucho más sencilla a presiones menores, por lo que los estados hinchados presentan una evolución con la presión mucho menor.

Por último, analizamos y mostramos en la Figura 4.9 los mapas de frecuencia de contactos pertenecientes a cada una de las cuatro proteínas obtenidos en condiciones extremas, presión muy alta y temperatura muy baja. Como ya mostrábamos en el Capítulo anterior, esto nos permite analizar la conformación hinchada ‘tipo’ de cada una de las proteínas, describiendo la conformación ‘ideal’ a la que nos llevaría nuestro modelo cuando es únicamente la presión la que actúa como agente desnaturizante, a temperatura ambiente.

De nuevo, los mapas que se muestran presentan diferencias significativas entre las dos familias de proteínas que se muestran, y que siguen la línea de la discusión desarrollada a lo largo de esta Sección. Las proteínas α , (a) y (c), muestran la presencia de un estado desnaturizado pero con características de un estado hinchado ‘perfecto’. Las hélices α están perfectamente formadas, como indican los contactos directos que las definen situados en los triángulos superiores de ambos mapas, mientras que los contactos entre hélices han pasado en su totalidad al estado hinchado, apareciendo con probabilidad cercana a la unidad en el triángulo inferior de los mapas. La baja temperatura hace que estos cálculos puedan considerarse desde un punto de vista práctico como una minimización de energía del sistema, y por eso la estructura resultante es demasiado estática, poco consistente con la idea de un estado desnaturizado mucho menos ordenado que

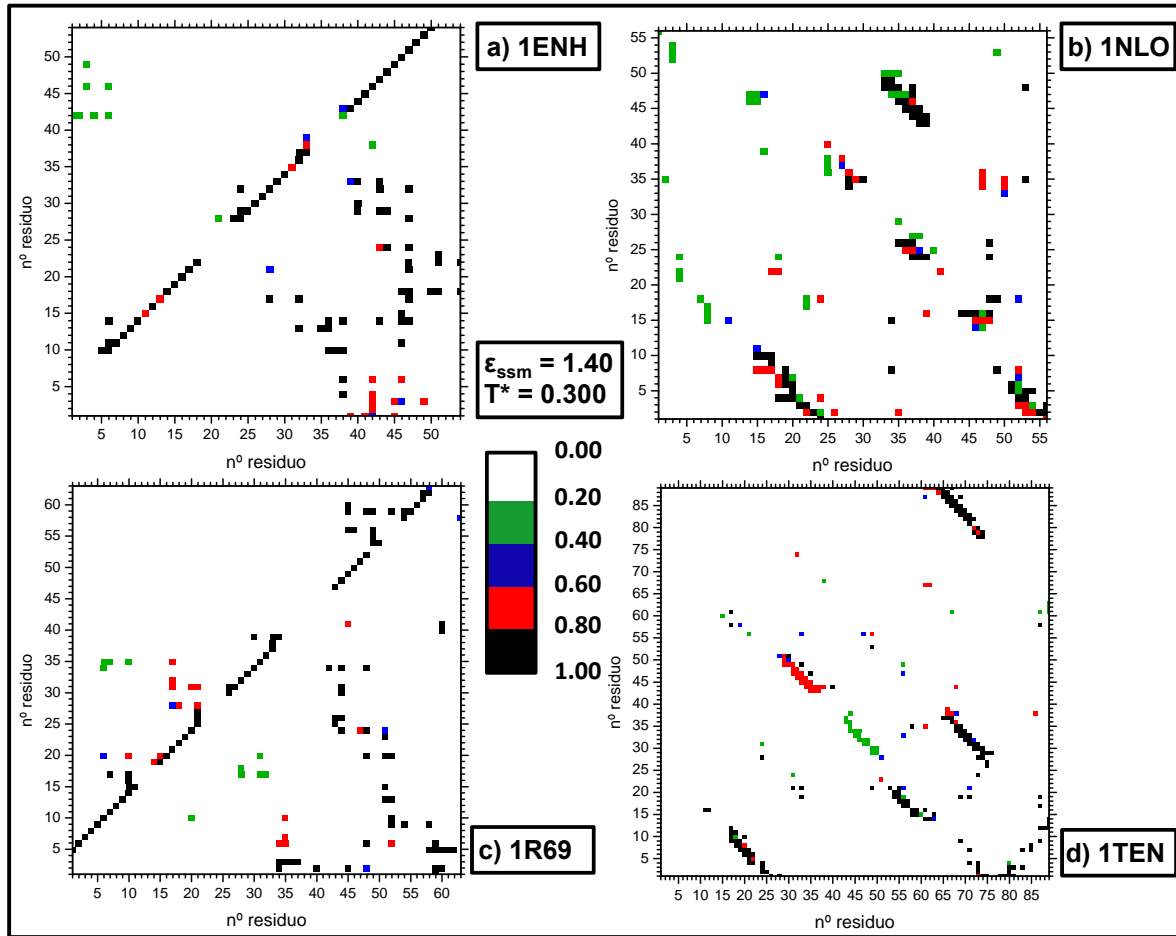


Figura 4.9: Mapas de frecuencia de contactos del estado hinchado en condiciones extremas. Se muestran los diferentes estados hinchados que hemos obtenido en condiciones de muy altas presiones y de muy bajas temperaturas.

el estado nativo de las proteínas correspondientes. En cualquier caso, permite definir con precisión las características del estado hinchado resultante de las interacciones del modelo.

Para las proteínas β , paneles (b) y (d) de la Figura 4.9, encontramos un estado no nativo con una estructura que mezcla contactos directos entre algunas hebras con contactos hinchados entre otros. Es interesante mencionar que la minimización de energía hace que algunas horquillas aparezcan preferentemente en el estado nativo, aunque la mayoría muestran claramente una preferencia considerable por el estado hinchado. La disminución de temperatura pone otra vez de manifiesto la frustración existente entre las interacciones de largo alcance y las de enlace de hidrógeno, encontrando diferentes

soluciones para cada caso concreto. Como ya hemos mencionado, como descripción del estado desnaturalizado por presión esta situación excesivamente estática a temperaturas muy bajas es posiblemente menos adecuada que la que se incluye en las gráficas de la Figura 4.7, siendo de nuevo relevantes los comentarios que hicimos en su discusión sobre las limitaciones del modelo de interacción en la descripción de este estado desnaturalizado.

Tras el análisis de cinco sistemas (incluyendo la proteína 2GB1 del Capítulo anterior) entre los que se encuentran dos proteínas con estructura secundaria todo β , dos con elementos de estructura secundaria todo α y una que mezcla ambos tipos de estructura secundaria, podemos afirmar que las proteínas con estructura β estabilizan con nuestro modelo un estado hinchado relativamente anómalo, que difiere del estado hinchado de naturaleza más realista de las proteínas α . En cualquier caso, con estas salvedades que ya hemos discutido, en todos los casos nos encontramos con un estado desnaturalizado por presión que sí puede considerarse un estado hinchado, en la línea de las observaciones experimentales. Además, concluimos que, como cabía esperar, cada familia de proteínas responde de manera diferente en su comportamiento estructural ante un aumento de presión, lo que puede considerarse un resultado favorable de nuestro modelo.

4.3.4. El diagramas de fases p - T para los cuatro sistemas

Uno de los resultados más relevantes de este trabajo, que ya presentamos en el Capítulo anterior, ha sido la posibilidad de reproducir el diagrama de fases p - T de la proteína 2GB1, aunque sea de forma cualitativa al introducir el parámetro ϵ_{ssm} en lugar de la presión en el eje de ordenadas. Por lo tanto, aprovechando las cuatro proteínas simuladas en este Capítulo, nos planteamos la posibilidad de realizar el mismo procedimiento para estos cuatro sistemas, y de esta manera comprobar si el modelo es capaz de reproducir el diagrama de fases elíptico esperado para cualquier proteína, como muestra la bibliografía^{40,42}. Además, el análisis de los diferentes diagramas de fases ϵ_{ssm} - T nos

permitirá seguir explorando posibles diferencias en los procesos de desnaturalización de las diferentes familias estructurales.

La Figura 4.10 (a) muestra en un único panel los diagramas de fases $\epsilon_{ssm}-T$ que hemos obtenido para las cuatro proteínas simuladas. El diagrama ha sido construido a partir de los puntos de cruce entre estados que hemos obtenido de los diagramas de poblaciones de las Figuras 4.4 y 4.5, como ya explicamos en la Sección 3.8. Lo primero que podemos observar al analizar el diagrama es cómo los cuatro sistemas muestran la forma elíptica que nos recuerda, de forma aproximada, a la región superior derecha del diagrama de fases $p-T$ genérico de las proteínas, como ya sucediese con la proteína 2GB1, cuyo diagrama de fases mostramos en la Figura 3.19. Por tanto, el modelo es capaz de reproducir el diagrama de fases de proteínas diferentes, independientemente del tipo de estructura secundaria que presenten. Además, confirma cómo la metodología puesta a punto durante el estudio de la proteína 2GB1 puede ser aplicado a otros sistemas.

Dado que las temperaturas de transición de cada una de las proteínas están relacionadas con su diferente tamaño y el número y distribución de sus contactos nativos, no tiene sentido en este contexto discutir las diferencias entre diferentes proteínas en cuanto a las temperaturas de la transición N-D a presiones bajas. Esta situación, además, puede llevar a error cuando se pretende dictaminar qué familia de proteínas es más estable frente al aumento de la presión a una temperatura determinada. Por esta razón, en el panel (b) de la misma Figura mostramos las líneas de fases N-D y N-H que se obtienen cuando se normalizan las temperaturas del mapa de poblaciones con respecto a la temperatura de transición de cada proteína a presión atmosférica. De esta forma obtenemos las cuatro líneas de fases superpuestas sobre un mismo punto. En esta representación es mucho más fácil de observar cómo a una temperatura equivalente para cada sistema, las proteínas α son más estables frente al aumento de la presión que las proteínas β , como se puede observar cuando nos situamos a una temperatura un 10 % inferior a la temperatura de transición de cada proteína, indicada por la línea discontinua de color

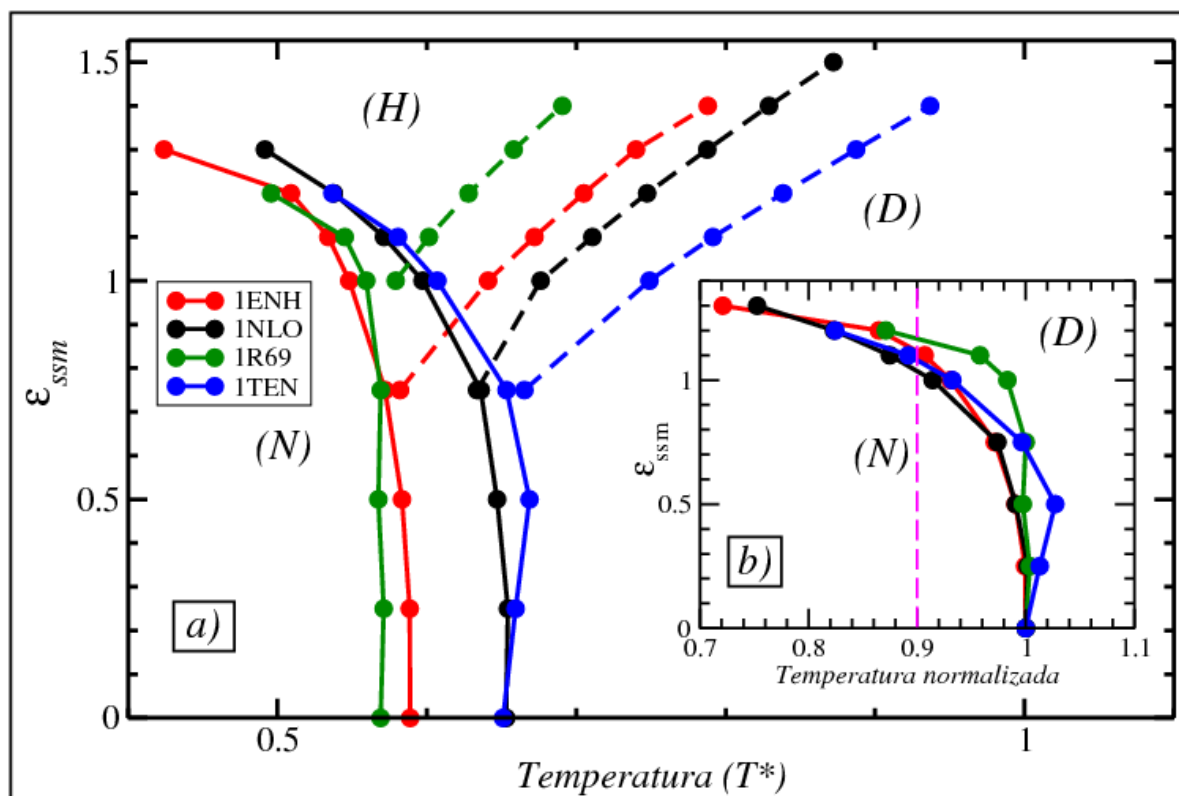


Figura 4.10: **Diagrama de fases $\epsilon_{ssm}-T$ para las cuatro proteínas.** a) Mostramos el diagrama de fases obtenido para los cuatro sistemas estudiados. Al igual que en la Figura 3.18, las líneas continuas muestran el tránsito entre el estado nativo y un estado desnaturalizado, mientras que las líneas discontinuas muestran un tránsito entre dos estados desnaturalizados. b) Comparación de los diagramas de fases normalizados para las distintas proteínas. La línea discontinua de color rosa nos muestra una temperatura un 10 % inferior a la temperatura de transición de cada uno de los sistemas.

rosa en el panel (b). Por lo tanto, la diferente estabilidad de las proteínas α y β frente a un aumento de presión también se refleja, como no podía ser de otra forma, en su diagrama de fases.

4.4. Resumen y conclusiones del Capítulo

A lo largo de este Capítulo hemos utilizado nuestro modelo de simulación para reproducir y estudiar el efecto de la presión en la desnaturalización de cuatro nuevos sistemas. Hemos seleccionado dos proteínas α y dos proteínas β para realizar las simulaciones con nuestro modelo de interacción. Así, además de obtener nuevas evidencias de cómo se comporta al aplicarlo a distintas proteínas, hemos podido realizar un estudio

de cómo la definición de las interacciones individuales en el modelo afecta a los dos principales elementos de estructura secundaria: las hélices α y las láminas β .

Hemos comprobado que el proceso de desnaturalización frente a la presión y la temperatura que obtenemos para estos cuatro sistemas es bastante similar al que encontramos para la proteína 2GB1. Encontramos tres regímenes de presión diferentes para cada sistema, con valores de ϵ_{ssm} que caracterizan a cada régimen de presión bastante parecidos entre unos sistemas y otros. Además, hemos demostrado la presencia de tres estados conformacionales diferentes en cada sistema: el estado nativo, el estado hinchado y el estado desnaturalizado. Hemos seleccionado los valores de radio de giro reducido y RMSD que caracterizan cada estado para calcular sus poblaciones en función de la temperatura y la presión del sistema. Además, hemos caracterizado los diferentes tipos de transiciones entre estados. En todos los casos observamos que el tránsito entre los estados nativo y desnaturalizado (térmico) es siempre más cooperativo que el tránsito entre el estado nativo y el hinchado (desnaturalizado por presión). A su vez, ambos resultan ser mucho más cooperativos que el tránsito entre el estado hinchado y el desnaturalizado térmico. También hemos hecho un análisis estructural de los estados hinchados obtenidos en cada sistema, y cómo van evolucionando con el aumento de la presión. Por último, hemos reproducido el diagrama de fases $\epsilon_{ssm}-T$ para las cuatro proteínas mediante la metodología que expusimos en el Capítulo anterior.

De los resultados que hemos obtenido a lo largo de este Capítulo podemos hacer dos lecturas principales. En primer lugar, hemos comprobado y demostrado que nuestro modelo de simulación es capaz de reproducir el proceso de desnaturalización por presión para proteínas con diferentes tamaños y tipos de estructura nativa, que muestran transiciones de dos estados a presión atmosférica. Por tanto, nos encontramos que el modelo desarrollado en esta Tesis, sencillo y de bajo coste computacional, proporciona resultados consistentes en sistemas estructuralmente diferentes, describiendo bastante bien de forma cualitativa el proceso de desnaturalización bajo presión hidrostática.

La segunda conclusión importante que podemos extraer de los resultados de este Capítulo es entender cómo la presión, pero también las características del modelo, afectan de distinta manera a las proteínas con estructura α y a las de estructura β . De forma general, hemos comprobado que la estructura β se muestra menos estable frente al aumento de presión en comparación con la estructura α . Con nuestro modelo, al alcanzar presiones altas las estructuras β evolucionan constantemente ante un aumento de presión, presentando un estado hinchado característico, en el que las interacciones hinchadas se van imponiendo a los enlaces de hidrógeno entre hebras β , resultando a presiones elevadas un estado hasta cierto punto poco realista, de mayor radio de giro y RMSD que el estado nativo, pero donde se produce un aumento de la distancia entre hebras que mantienen su estructura secundaria. Por su parte las estructuras α estabilizan, dentro de un régimen de presiones elevadas, un estado hinchado mejor definido, con las hélices bien formadas y una estructura terciaria hinchada, más plausible con la idea de un estado hinchado por inyección de moléculas de agua como el que se describe en la bibliografía para los estados hinchados de una proteína.

Finalmente, el diagrama de fases $\epsilon_{ssm}-T$ de las cuatro proteínas nos demuestra que nuestro modelo sirve para describir de forma adecuada las características de los diferentes tránsitos existentes en el proceso de plegamiento/desnaturalización de diferentes proteínas en función de la temperatura y la presión, sea cual sea la estructura secundaria de las proteínas consideradas.

En resumen, el modelo desarrollado no sólo reproduce resultados satisfactorios para ambas familias estructurales, sino que además identifica estados hinchados con diferentes propiedades según se trate de una proteína α o de una β , a pesar de que no introducir a priori ninguna condición en la simulación que permita distinguir ambos casos. Es el propio mapa de contactos de la proteína, junto con la implementación de los contactos de enlace de hidrógeno, el responsable de las diferencias observadas entre proteínas con diferentes estructuras nativas.

Capítulo 5

Comparación con resultados experimentales

En los Capítulos anteriores de esta memoria hemos evaluado nuestro modelo, aplicándolo a proteínas con diferentes estructuras. Hemos comprobado su validez mediante la obtención de estados desnaturalizados por presión, o estados hinchados, con características que dependen de la estructura nativa de partida. De esta manera, hemos comprobado que las características globales que se describen en la bibliografía para los estados desnaturalizados por presión se pueden reproducir, al menos de forma cualitativa, con nuestro modelo de simulación. Sin embargo, no hemos intentado ir más allá, puesto que apenas existen en la bibliografía datos experimentales concretos que podamos comparar de forma más directa con nuestros resultados de simulación. En este Capítulo intentamos una comparación de estas características.

5.1. Motivación del Capítulo

En el último Capítulo de resultados de esta Tesis, pretendemos estudiar de nuevo el proceso de desnaturalización de una proteína bajo presión hidrostática, para posteriormente comparar los resultados de nuestro modelo con medidas experimentales para

esa misma proteína. Existen multitud de proteínas cuyo plegamiento ha sido ampliamente estudiado desde diferentes puntos de vista^{181,182}. Sin embargo, no son tantas las proteínas para las cuales podemos encontrar resultados de su desnaturalización bajo presión hidrostática. Como ya comentamos en la Sección 1.4, son principalmente motivos técnicos los que han impedido hasta fechas muy recientes realizar un estudio del plegamiento de proteínas bajo unas condiciones variables de presión y temperatura. Por tanto, no resulta tan sencillo comparar directamente los resultados de nuestra simulación con resultados obtenidos de manera experimental en el laboratorio.

Además, debemos tener en cuenta que la mayoría de los estudios en los que se somete una proteína a un aumento de presión hidrostática no pretenden hacer un análisis detallado del proceso de desnaturalización, sino que lo más habitual es aprovechar un aumento moderado de la presión en el medio para estudiar otras propiedades, por ejemplo el efecto de las cavidades en el interior de la proteína al aumentar la presión⁶² o efectos de agregación¹⁸³. A todo esto hay que añadir las dificultades que encontramos al comparar los resultados de nuestro modelo *coarse-grained* con estudios de proteínas reales, puesto que muchos de los trabajos presentes en la bibliografía hacen referencia a cómo varían detalles estructurales a nivel atómico de la proteína a medida que aumenta la presión del medio^{184,185}. Por lo tanto, resulta realmente difícil obtener información global acerca de cómo evoluciona el estado nativo, en su proceso de desnaturalización frente a la presión del medio, que pueda ser comparada con el tipo de resultados que obtenemos al aplicar nuestro modelo.

Ante esta situación, nos planteamos como primer objetivo para este Capítulo encontrar en la bibliografía resultados experimentales de alguna proteína cuya desnaturalización a altas presiones haya sido estudiada, y que proporcionen información sobre características del proceso que puedan ser calculadas a través de nuestras simulaciones. Este Capítulo resulta ser importante en el proyecto de esta Tesis por dos motivos principales. Por un lado, para comprobar cómo nuestros resultados se confirman en mayor

o menor medida, cuando en vez de realizar un experimento virtual realizamos un experimento con proteínas reales. Recordemos que, además, hasta ahora hemos mostrado siempre resultados fundamentalmente cualitativos, y por tanto la comparación de nuestros resultados con datos experimentales logrará que nuestro modelo pueda adoptar una vertiente más cuantitativa. Además, si se cumple este primer objetivo, y por tanto queda demostrado que nuestro modelo es capaz de reproducir los resultados experimentales, alcanzaremos una segunda conclusión importante para el modelo desarrollado en esta Tesis: los resultados presentados hasta el momento pueden tomarse como predictivos de lo que realmente sucedería si estudiásemos experimentalmente la desnaturalización de las proteínas 2GB1, 1ENH, 1R69, 1NLO y 1TEN bajo un aumento de la presión hidrostática, ya que para las mismas no hay demasiada información en este sentido. De esta manera, conseguiríamos además uno de los objetivos principales de esta Tesis, que consiste en poder seguir aumentando el conocimiento básico de los procesos de desnaturalización mediados por presión para diferentes proteínas.

Creemos, por tanto, que contrastar los resultados obtenidos de nuestro modelo con resultados reales puede ser el paso definitivo para que el modelo que hemos ido describiendo a lo largo de la Tesis pueda ser de utilidad contrastada para la comunidad científica.

5.2. La proteína SNasa

La proteína seleccionada para este estudio ha sido la proteína *nucleasa de estafilococo*, que se suele representar en inglés por el acrónimo *SNase*, y al que nos referiremos en este capítulo como SNasa. Esta proteína, presente en las bacterias de la familia de los estafilococos, es una enzima perteneciente a la familia de las hidrolasas, cuya funcionalidad biológica consiste en la ruptura de los enlaces fosfodiéster de la cadena de los ácidos nucleicos. Al tratarse de una proteína muy estudiada, hemos identificado más de 200 estructuras para esta proteína y sus mutantes en el PDB¹³⁷. Nosotros hemos decidido

utilizar la estructura cuyo código PDB es 1STN¹⁸⁶, al ser una de las estructuras más utilizadas de esta proteína cuando se realizan estudios por simulación molecular con detalle atómico^{187–189}. SNasa presenta dos diferencias significativas con respecto a las otras proteínas que hemos estudiado en esta Tesis. Se trata de una proteína de un tamaño significativamente mayor, 136 aminoácidos, que además presenta un plegamiento de tres estados a presión atmosférica^{28,190,191}, a diferencia del resto de proteínas analizadas hasta el momento, que presentan siempre un plegamiento de dos estados. Si hemos seleccionado entonces esta proteína es porque se trata de una de las pocas proteínas con la que más se ha trabajado en condiciones de alta presión, y por tanto en principio parece la proteína idónea para ser simulada con nuestro modelo. Aunque sea de un tamaño algo mayor del que solemos utilizar, el tiempo requerido para cada una de las simulaciones, consiste en un barrido de temperaturas para cada valor de presión, es aproximadamente de una semana (tiempo de ordenador), por lo que entra dentro de las posibilidades de nuestro equipos informáticos, al permitirnos realizar la simulación completa del sistema dentro de unos tiempos razonables. En la Figura 5.1 presentamos el mapa de contactos nativos para esta proteína. Como podemos apreciar, se trata de una proteína α/β con 3 hélices α y 6 hebras β . En total presenta 585 contactos, de los cuales 267 son de corto alcance y 318 son de largo alcance, 56 de los cuales son contactos de enlace de hidrógeno. El radio de giro de esta estructura PDB calculado sobre los carbonos α es de 14.0 Å.

A pesar de ser posiblemente la proteína con la que se ha realizado un mayor número de experimentos a alta presión, como hemos mencionado antes, muchos de estos estudios no perseguían estudiar el proceso de desnaturalización. Por el contrario, se han centrado en otras características que se dan en las proteínas con el aumento de la presión. Entre ellas se encuentran, para esta proteína concreta, posibles cambios en las propiedades de solvatación¹⁸⁹, cambios de volumen asociados al aumento de la presión¹⁹², la importancia de las cavidades en el interior de la estructura globular¹⁹³ o la desestabilización de agregados formados por un mal plegamiento¹⁹⁴. Por tanto, los

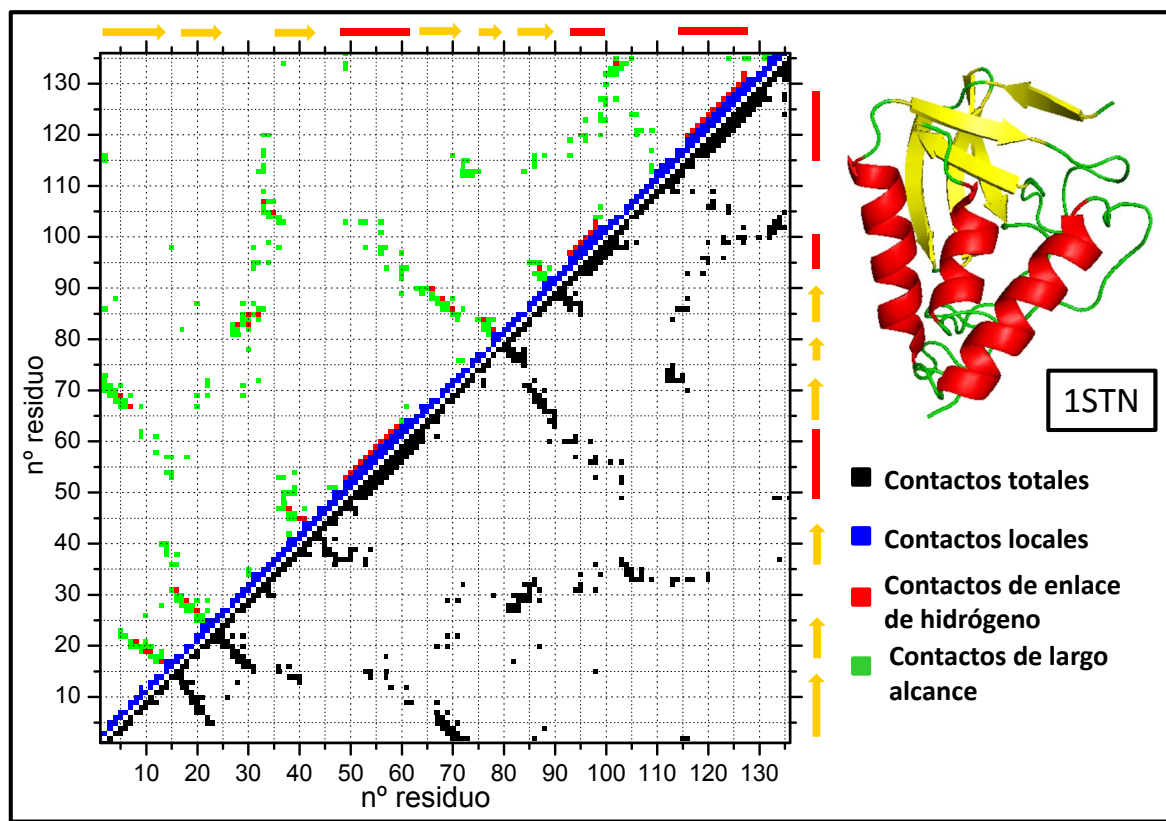


Figura 5.1: **Mapa de contactos nativos de la proteína 1STN.** El triángulo inferior muestra el mapa de contactos nativos, el triángulo superior los mismos contactos clasificados en función de su tipo. Las flechas naranjas en los ejes muestran las posiciones de las hebras β , mientras que las barras rojas marcan la posición de las hélices α .

trabajos con los que podemos comparar los resultados obtenidos son bastantes menos de los deseados. Aun así, son más de los que se encuentran para la mayoría de las proteínas.

5.3. El plegamiento a presión atmosférica

El primer reto que se nos presenta al seleccionar esta proteína es observar si nuestro modelo es capaz de reproducir un proceso de más de dos estados a presión atmosférica. En caso afirmativo nos serviría para poder validar el modelo no sólo para proteínas que presentan plegamiento de dos estados, sino también para proteínas que pliegan en tres o más estados. Recordemos que muchas de las proteínas de tamaño medio o grande presentan un plegamiento de más de dos estados²⁷.

Para saber si nuestro modelo es capaz de estabilizar a presión atmosférica un estado intermedio entre el estado nativo (N) y el estado desnaturalizado térmico (D), realizamos un estudio de las configuraciones presentes a una temperatura cercana al máximo de la curva de capacidad calorífica cuando $\epsilon_{ssm} = 0.00$. Para este valor de ϵ_{ssm} , la curva de capacidad calorífica (que podemos ver en la Figura 5.3, comparada con el resto de curvas medidas a distintas presiones) muestra un único pico estrecho cuyo máximo se sitúa a $T^* = 0.603$. En la Figura 5.2 mostramos el histograma de energía y el perfil de energía de Helmholtz que obtenemos a esta temperatura de transición, paneles (b) y (e), respectivamente, junto con los que encontramos a una temperatura inferior a la del máximo de la curva de capacidad calorífica, $T^* = 0.599$, paneles (a) y (d); y a una temperatura superior a la del máximo de la curva de capacidad calorífica, $T^* = 0.607$, paneles (c) y (f).

En estas representaciones observamos cómo, además del estado N de menor energía, centrado en un valor de energía de aproximadamente -450 unidades reducidas (u.r.), y el estado D de mayor energía, aproximadamente a -150 u.r., se aprecia la presencia de un intermedio centrado en aproximadamente -370 u.r. Podemos apreciar que se trata de un intermedio cuya población es muy inferior a la de los estados N y D, a tenor de la pequeña frecuencia que muestran para este estado los histogramas de energía que observamos a cualquiera de las tres temperaturas recogidas en la Figura. Además, en el perfil de energía de Helmholtz podemos observar cómo el intermedio presenta una barrera pequeña que lo separa del estado nativo y una energía bastante parecida a este, lo que explica la presencia de únicamente un pico en la curva de capacidad calorífica. En la Figura 5.2 también podemos observar la presencia de un segundo intermedio, centrado en una energía de aproximadamente -270 u.r., pero cuya población en los histogramas es prácticamente inexistente, ya que en los perfiles de energía de Helmholtz aparece a valores sensiblemente superiores a los de los estados más estables a cada temperatura.

Puesto que la existencia de un estado intermedio podría afectar a nuestro posterior

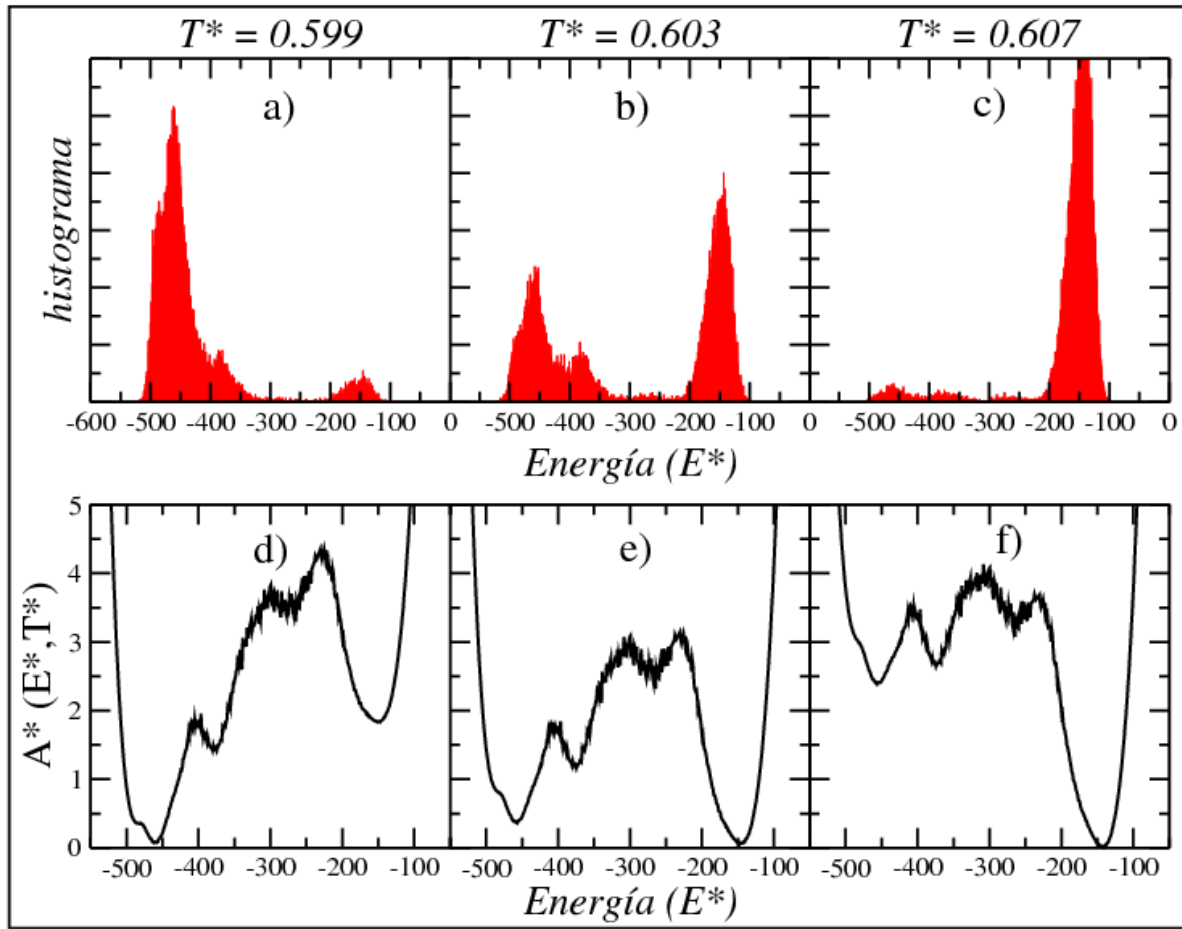


Figura 5.2: **Presencia de un intermedio entre el estado nativo y el estado desnaturalizado a presión atmosférica.** Se presentan los histogramas de energía en la fila superior, y los perfiles de energía de Helmholtz en la fila inferior, que hemos calculado para las simulaciones a presión atmosférica, a una temperatura inferior a T_m , paneles (a) y (d); a la temperatura de transición T_m , paneles (b) y (e); y a una temperatura superior a T_m , paneles (c) y (f).

análisis del estado hinchado a mediada que aumenta la presión del sistema, decidimos realizar un estudio de la evolución del estado intermedio a medida que aumenta la presión del sistema. Pudimos contemplar (datos no mostrados) cómo a medida que aumenta el valor de ϵ_{ssm} existe una disminución considerable en la población del estado intermedio, hasta el punto que cuando $\epsilon_{ssm} = 0.50$ la población del estado intermedio es prácticamente nula en comparación con la de los estados nativo y desnaturalizado. Por esta razón, podemos concluir que el estado intermedio que hemos observado no interfiere en el muestreo del sistema cuando aumentemos el valor del parámetro ϵ_{ssm} .

De esta forma, queda demostrado cómo el modelo que hemos desarrollado es capaz

de reproducir plegamientos de más de dos estados, como describe la bibliografía para esta proteína^{28,190,191}. El intermedio detectado presenta además valores de radio de giro y RMSD intermedios entre los dos estados de referencia. Este resultado nos permite demostrar cómo el modelo desarrollado no es únicamente válido para proteínas que pliegan en procesos de dos estados, sino que resulta ser también efectivo para la simulación de proteínas con procesos de desnaturalización más complejos, lo que sin duda sirve para seguir reforzando la validez del modelo desarrollado.

5.4. Aparición del estado hinchado con el aumento de presión.

Una vez hemos comprobado que el modelo se ajusta a los resultados experimentales, al menos a presión atmosférica, empezamos a analizar su evolución a medida que aumenta la presión hidrostática del medio. Para ello calculamos las curvas de capacidad calorífica para cada una de las condiciones de simulación. En la Figura 5.3 podemos observar la evolución de la curva de capacidad calorífica a medida que aumenta el valor del parámetro ϵ_{ssm} .

En la Figura se observa una vez más la evolución de las características de la curva a medida que aumenta el valor de ϵ_{ssm} . Apoyándonos además en los resultados que hemos ido presentando hasta el momento para el resto de proteínas estudiadas en este trabajo, podemos apreciar el descenso en la cooperatividad del tránsito a medida que aumenta la presión. También podemos intuir de nuevo la presencia de tres regímenes de presión diferentes en función de la forma que adoptan las curvas, siendo a priori 0.75/1.00 y 1.20/1.30 los valores de ϵ_{ssm} que actúan de frontera entre los diferentes regímenes de presión. La mayor diferencia que podemos observar entre esta representación de las curvas de capacidad calorífica y las que hemos presentado para el resto de proteínas en los capítulos anteriores es que para valores pequeños de ϵ_{ssm} existe un desplazamiento

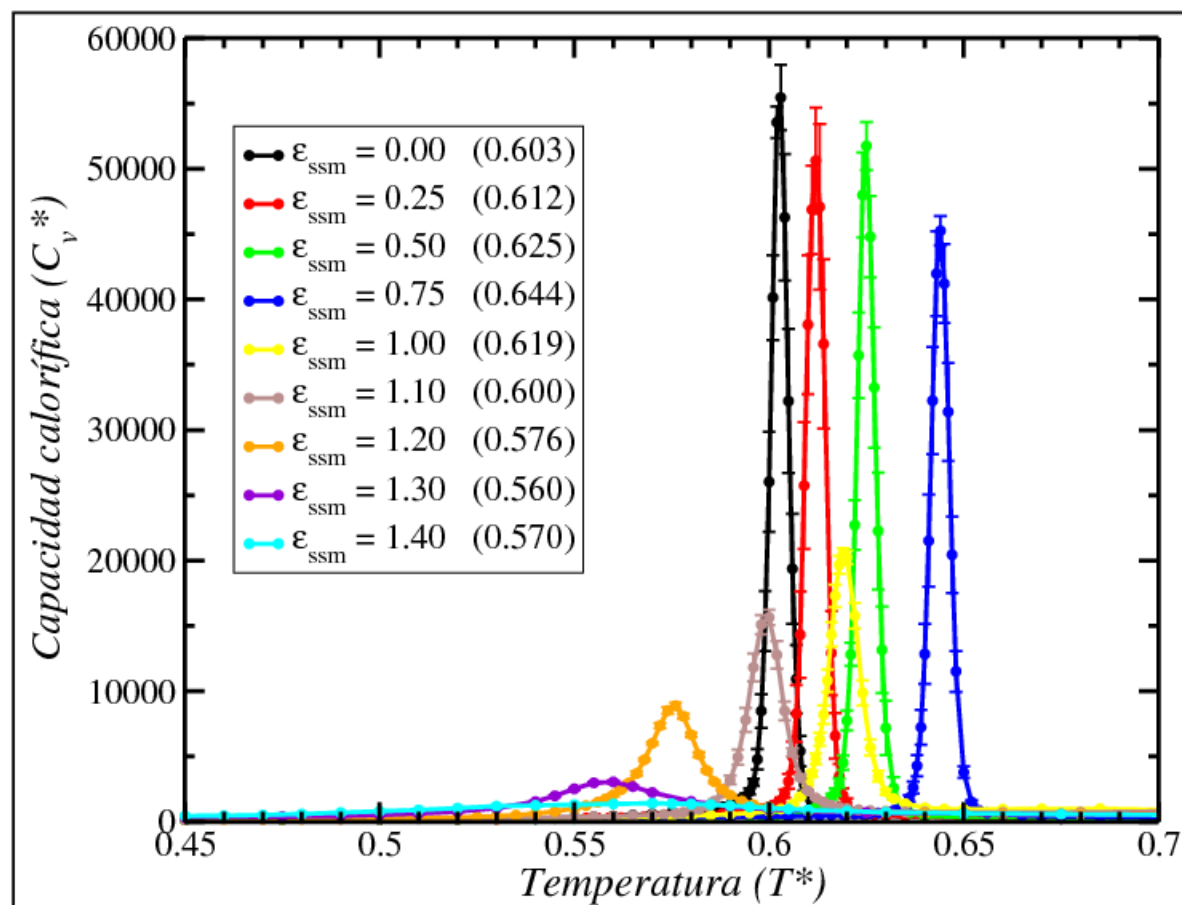


Figura 5.3: **Evolución de las curvas de capacidad calorífica con la presión.** Se presentan las curvas de capacidad calorífica calculadas para cada valor de ϵ_{ssm} . Entre paréntesis se muestra la temperatura correspondiente al máximo de la curva.

neto de T_m hacia valores mayores de temperatura. Recordemos que este comportamiento para valores bajos de ϵ_{ssm} sólo lo habíamos observado en el Capítulo 4 para la proteína 1TEN, que al igual que 1STN se trata de una proteína de mayor tamaño que el resto de las proteínas simuladas en esta Tesis. Por lo tanto, es posible que esta característica de una ligera estabilización del estado nativo con la presión para valores relativamente bajos de esta se encuentre relacionada con el tamaño de la cadena, y por esa razón todas las demás proteínas de pequeño tamaño que hemos simulado no presentan este comportamiento. En cualquier caso, a la vista de las curvas de capacidad calorífica podemos afirmar que esta proteína presenta un proceso de desnaturalización por presión similar al que hemos observado hasta el momento para el resto de proteínas estudiadas.

Como hemos comentado previamente, existen en la bibliografía diversos trabajos experimentales para SNasa que muestran la existencia de un estado desnaturalizado por efecto de la presión que difiere notablemente del estado desnaturalizado por efecto de la temperatura^{45,188}. En dichos trabajos se caracterizan estos estados desnaturalizados a través de los valores del radio de giro promedio que presenta la proteína a medida que se desnaturaliza, demostrando que un aumento de la presión nos lleva a la aparición de un estado con radio de giro inferior al que presenta el estado desnaturalizado térmicamente. En la Figura 5.4 (a), tomada del trabajo de Paliwal y col.¹⁸⁸, podemos observar la variación del radio de giro promedio de la cadena polipeptídica con la presión hidrostática del medio a temperatura ambiente. Queda patente cómo existe un aumento del radio de giro de la estructura a medida que aumenta la presión, estabilizándose una situación con radio de giro de aproximadamente 35 Å, que corresponde al estado hinchado de la proteína 1STN. Este estado presenta un radio de giro inferior al radio de giro de la estructura desnaturalizada por efecto de la temperatura, como podemos ver en el panel (b) de la Figura 5.4. En esta Figura, tomada del trabajo de Panick y col.⁴⁵, podemos observar cómo a 80 °C y presión atmosférica se aprecia la presencia de un estado con radio de giro promedio de aproximadamente 65 Å, muy superior al mostrado para el estado hinchado. Un radio de giro tan grande es característico del estado desnaturalizado por temperatura. Por tanto, parece que el valor del radio de giro es una buena propiedad para discernir experimentalmente entre los estados hinchado y desnaturalizado de esta proteína.

Nosotros nos propusimos realizar del mismo modo un análisis del valor del radio de giro obtenido de nuestras simulaciones en función del aumento de la temperatura y de la presión, o en nuestro caso del parámetro ϵ_{ssm} . En la Figura 5.5 podemos ver la evolución del radio de giro promedio con la temperatura para cada una de las simulaciones realizadas. Recordemos que, como mostramos en la ecuación 2.10, los resultados que presentamos en esta Tesis para el radio de giro corresponden a un valor normalizado con

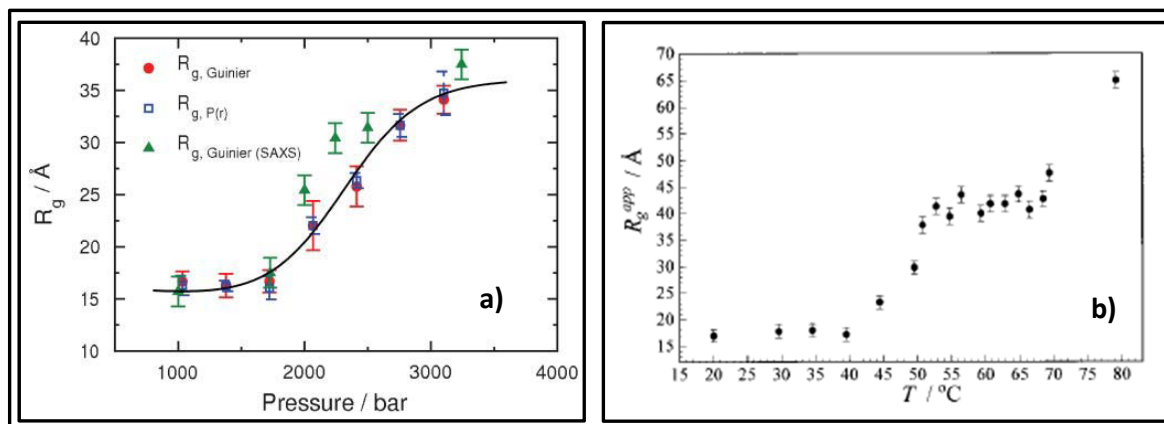


Figura 5.4: **Evolución del radio de giro con la presión y la temperatura para la proteína SNasa.** a) Medida experimental de la evolución del radio de giro con la presión a temperatura ambiente, tomada del trabajo de Paliwal y col.¹⁸⁸. b) Medida experimental de la evolución del radio de giro con la temperatura a presión atmosférica, tomada del trabajo de Panick y col.⁴⁵.

respecto al radio de giro de la proteína nativa. Se pueden observar tres comportamientos bien diferenciados en función del valor que toma ϵ_{ssm} . Para valores de $\epsilon_{ssm} < 1.00$ existe un salto en el valor del radio de giro reducido que pasa de aproximadamente 1.00 hasta un valor superior a 2.2. Este cambio en el radio de giro promedio es muy brusco y se da siempre a una temperatura que coincide con el valor de T_m que habíamos obtenido para cada una de las simulaciones, ver Figura 5.3. Este cambio muestra el paso desde el estado N, con radio de giro reducido cercano a 1.00, hasta el estado D con radio de giro reducido > 2.2 .

Para valores de ϵ_{ssm} entre 1.00 y 1.20 notamos cómo existe, alrededor de T_m , un primer cambio en el radio de giro reducido mucho menos acusado que el de presiones más bajas; además, podemos apreciar cómo durante un intervalo estrecho de temperaturas se estabiliza un estado con un radio de giro reducido cercano a 1.1. Un valor de 1.1 en el radio de giro reducido es característico de un estado desnaturalizado bastante compacto, como el que se estabiliza por efecto de la presión, y por tanto debe corresponder al estado H de esta proteína. Si la temperatura sigue aumentando, dicho estado se desestabiliza evolucionando de forma continua hasta estructuras con un radio de giro reducido cercano a 2.0, característico del estado D.

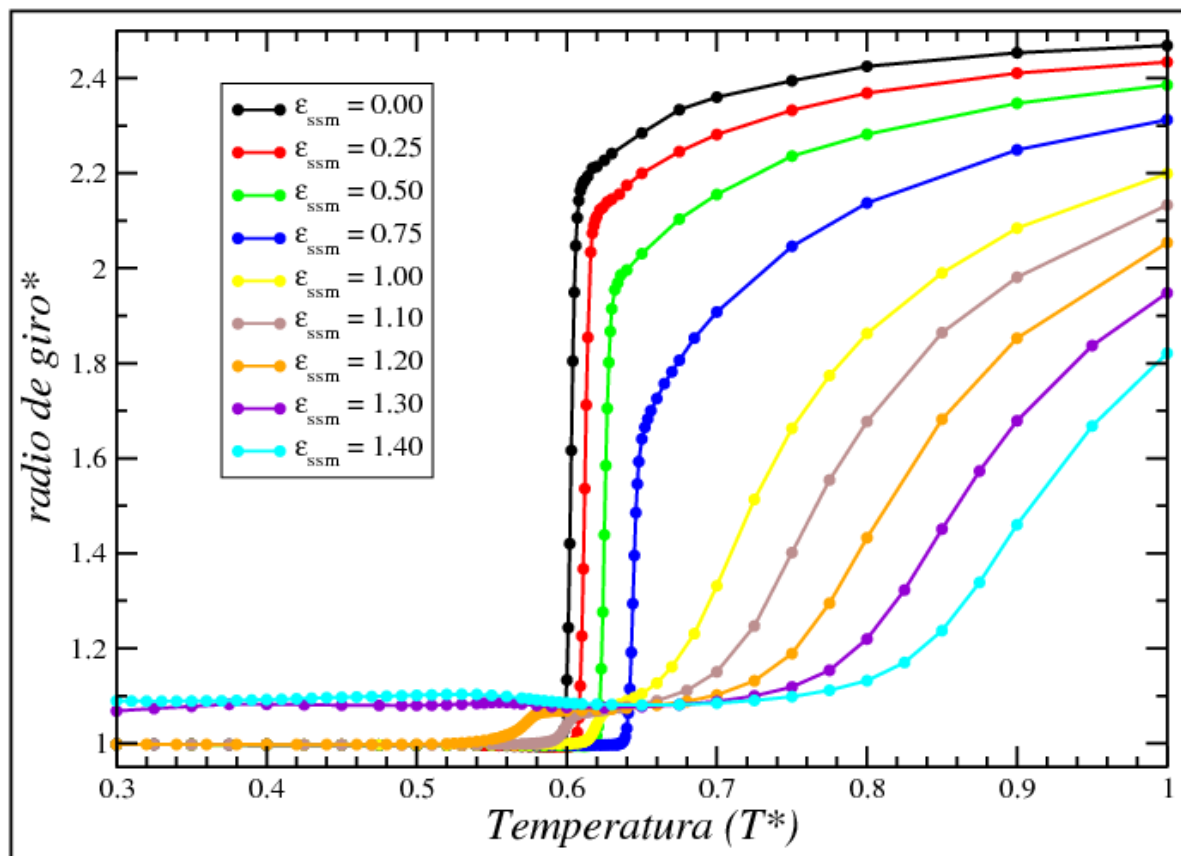


Figura 5.5: **Evolución del radio de giro con la temperatura, según nuestro modelo.** Evolución del radio de giro (reducido con el valor de la estructura nativa) con la temperatura para los distintos valores de ϵ_{ssm} .

Cuando el valor de ϵ_{ssm} es igual a 1.2 o superior, lo que observamos es que desde las primeras temperaturas de nuestra simulación se estabiliza el estado hinchado de la proteína, con radio de giro reducido cercano a 1.1, que permanece estable hasta valores elevados de temperatura, donde comienza a desestabilizarse al igual que sucedía en el caso anterior.

Por otro lado, la representación del radio de giro promedio de las conformaciones obtenidas para esta proteína a una misma temperatura pero diferentes valores de ϵ_{ssm} debería mostrarnos la evolución existente entre los diferentes estados, al igual que sucedía con los resultados experimentales de la Figura 5.4 (a). Además, nos debería permitir relacionar los valores del parámetro ϵ_{ssm} con una presión real. En la Figura 5.6 observamos la evolución del radio de giro medio con respecto a ϵ_{ssm} a dos temperatu-

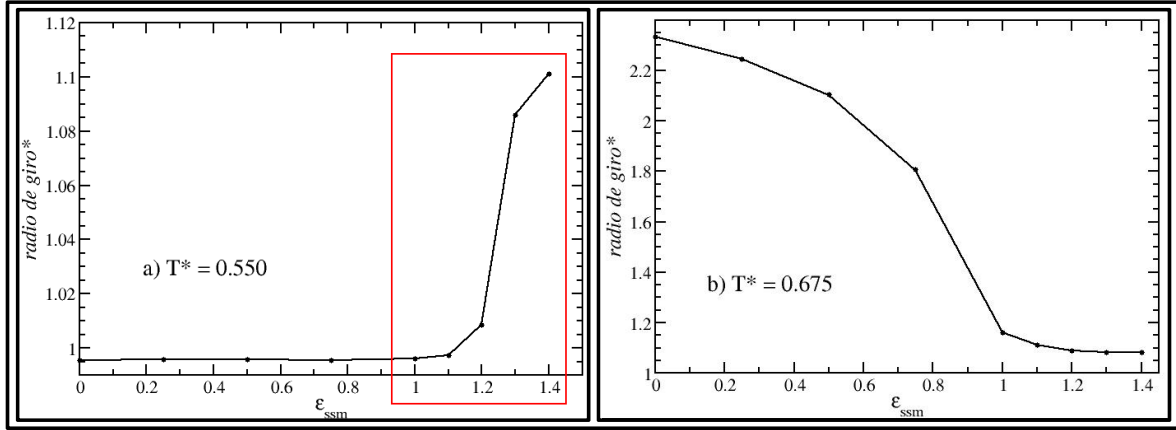


Figura 5.6: **Evolución del radio de giro con la temperatura, según nuestro modelo.** Se muestra la evolución del radio de giro reducido promedio en función del valor que toma ϵ_{ssm} a una temperatura constante. a) $T^* = 0.550$. b) $T^* = 0.675$.

ras diferentes. En la primera de ellas, panel (a) mostramos la representación cuando $T^* = 0.550$, temperatura inferior al máximo de la curva de capacidad calorífica para la mayoría de las presiones simuladas. En la Figura podemos notar cómo para valores de ϵ_{ssm} inferiores a la unidad apenas existe evolución del radio de giro, ya que nos encontramos en un régimen de bajas presiones, donde el aumento de ϵ_{ssm} apenas influye en la estructura nativa. Sin embargo, cuando el valor de $\epsilon_{ssm} > 1.00$, observamos cómo existe un aumento apreciable del radio de giro a medida que aumenta la presión. Nos encontramos por tanto en unas condiciones en las que el efecto de la presión es suficiente como para inducir cambios significativos en el sistema. Si observamos la parte derecha de la Figura 5.6 (a), zona de mayores presiones, señalada con un recuadro rojo en la gráfica, encontramos que existe una cierta similitud entre la forma de la curva de los resultados de simulación y la Figura 5.4 (a), que representa datos experimentales para la misma proteína. Para ambas gráficas notamos cómo a valores altos de presión existe un aumento del radio de giro. En los resultados de nuestra simulación podemos notar la estabilización del estado hinchado con radio de giro reducido cercano a 1.1.

Cuando realizamos el mismo tipo de análisis a una temperatura relativamente alta, $T^* = 0.675$, observamos un comportamiento diferente, Figura 5.4 (b). Podemos notar cómo el radio de giro decrece a medida que aumenta el valor de ϵ_{ssm} . En este

caso, para valores bajos de ϵ_{ssm} se estabiliza el estado D, con radio de giro reducido superior a 2.0. Para esta temperatura hemos sobrepasado el máximo de la curva de capacidad calorífica y por tanto el estado que se estabiliza a presión atmosférica es el estado desnaturalizado por efecto de la temperatura. A medida que aumenta la presión del sistema, observamos cómo el estado D va disminuyendo progresivamente su radio de giro, hasta que $\epsilon_{ssm} = 1.0$; a partir de este valor, el sistema estabiliza al estado H con radio de giro reducido cercano a 1.1. Queda así demostrada la evolución en nuestros resultados de simulación desde el estado D hasta el estado H a medida que aumenta la presión del sistema, un comportamiento característico de temperaturas elevadas.

Podemos notar, por tanto, que a partir de los resultados que se muestran en las Figuras 5.5 y 5.6 demostramos para nuestros resultados de SNasa la existencia de tres estados diferentes, nativo, hinchado y desnaturalizado, y de los tres regímenes de presión que hemos ido identificado para todas nuestras proteínas: bajas presiones, presiones moderadas y altas presiones. Además, gracias a los resultados experimentales que se muestran en la Figura 5.4 (a), podemos relacionar los valores de ϵ_{ssm} con una presión determinada, como mostramos en la Tabla 5.1. Esta relación entre el valor de ϵ_{ssm} y la presión del sistema es válida para esta proteína. Sin embargo, puesto que para todas las proteínas estudiadas hasta el momento los intervalos han resultado ser similares (Tabla 4.3), y además sabemos que los intervalos de presión en los que desnaturalizan las proteínas no son excesivamente amplios^{40,43}, podemos, siempre pendiente de confirmaciones posteriores, relacionar de forma cuantitativa los valores de ϵ_{ssm} y la presión a la que se encuentra el sistema.

Por último cabe destacar que, pese a que habría sido muy interesante comparar de forma cuantitativa el valor del radio de giro promedio que se presenta en los trabajos experimentales citados y el radio de giro promedio que resulta de nuestro análisis, dicha comparación no es directamente posible. El radio de giro que se presenta en estos dos trabajos experimentales resulta ser en realidad un radio hidrodinámico ya que, expe-

intervalo de presiones (bar)	intervalo de ϵ_{ssm}
0 - 1400	(0.00 - 1.00)
1400 - 3500	(1.00 - 1.40)
>3500	(>1.40)

Tabla 5.1: Relación cuantitativa entre el valor de ϵ_{ssm} y la presión del sistema.

rimentalmente, no es factible medir el radio de giro de una molécula con una técnica dinámica como la utilizada en los trabajos mencionados. La medida del radio de giro real se puede realizar únicamente con alguna técnica estática¹⁹⁵. Al mismo tiempo, la comparación entre el radio de giro verdadero y el radio hidrodinámico no es factible ni siquiera de forma cualitativa, ya que la relación cuantitativa entre ambas magnitudes, radio de giro y radio hidrodinámico, no es una constante sino que depende de la conformación que adopta la molécula, y por tanto es un parámetro que varía en función del estado que estamos estudiando en cada momento¹⁹⁶. Esto explica por qué los valores cuantitativos de radio de giro que se presentan experimentalmente para los estados hinchado y desnaturalizado no coinciden con los que obtenemos en nuestra simulación, y no podemos relacionarlos directamente con los valores de nuestro radio de giro obtenido de las simulaciones.

5.5. Comparación de los diagramas de fases.

Otro de los resultados que ha sido medido experimentalmente para esta proteína, y que podemos comparar de forma directa con nuestro modelo, es el diagrama de fases p - T ^{41,197}. En los trabajos experimentales se presentan diferentes diagramas de fases para la proteína SNasa medidos en condiciones distintas. En la Figura 5.7 mostramos estos diagramas, tomados del trabajo de Panick y col.⁴¹, donde se determina el diagrama de fases hasta de tres maneras diferentes: mediante medidas de fluorescencia, por medidas de absorción en el infrarrojo (IR) y mediante medidas de difracción de rayos X a ángulos bajos, SAXS.

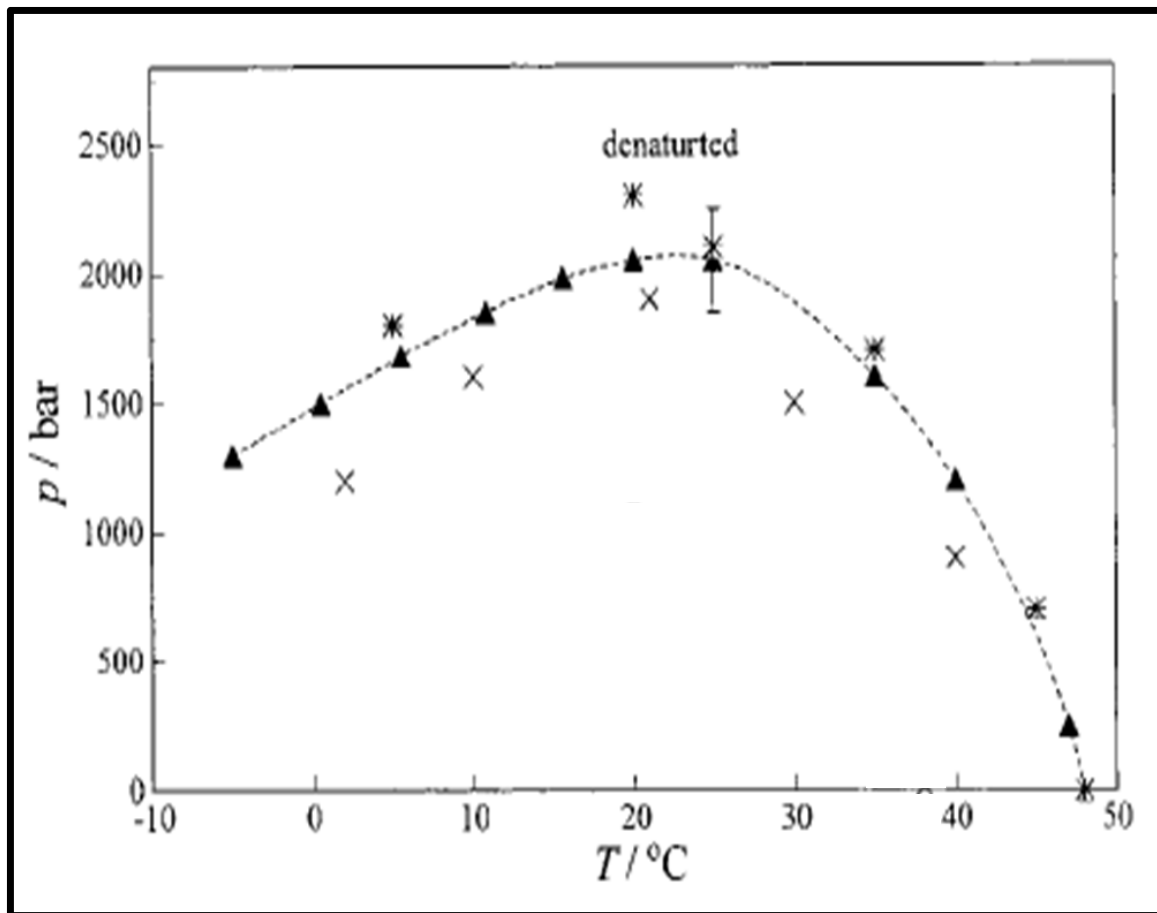


Figura 5.7: **Diagramas de fases p - T para la proteína SNasa.** Se presentan diferentes diagramas de fases experimentales para la proteína SNasa obtenidos por técnicas diferentes: fluorescencia (\times), absorción FTIR (\triangle) y SAXS (*).

Podemos observar cómo se obtienen diferentes diagramas de fases en función del método empleado para su determinación, aunque todos los diagramas presentados tienen una forma general bastante similar.

Para realizar nuestro diagrama de fases a partir de los resultados de simulación operamos de la misma manera en que lo hemos hecho con el resto de proteínas. Clasificamos cada una de las conformaciones registradas de la proteína como estado nativo, estado hinchado y estado desnaturalizado en función de sus valores de radio de giro y RMSD (recordamos que el intermedio que aparece a presión atmosférica no se detecta a presiones superiores). Posteriormente, realizamos el mapa de poblaciones para los tres estados en función de la temperatura para cada valor de ϵ_{ssm} , y a partir de los

	Estado nativo	Estado hinchado	Estado desnaturalizado
R_g^*	<1.025	$1.025 - 1.15$	>1.15
RMSD (Å)	<2.0	$2.0 - 12.5$	>12.5

Tabla 5.2: Intervalos de las propiedades estructurales que permiten identificar las conformaciones pertenecientes a los diferentes estados en la proteína SNasa.

puntos de corte entre las curvas correspondientes a cada estado obtenemos los puntos característicos de nuestro diagrama de fases. En la Tabla 5.2 mostramos los parámetros que identifican a cada uno de los estados para esta proteína. El diagrama de fases que hemos obtenido para 1STN se muestra en la Figura 5.8. Podemos apreciar cómo nuestro diagrama de fases, $\epsilon_{ssm}-T$, no tiene exactamente la misma forma que los que presenta la bibliografía experimental para esta proteína, Figura 5.7. Sin embargo, sí muestra similitudes, y por tanto parece interesante hacer una comparación entre los resultados hallados en la bibliografía y los que obtenemos al realizar nuestra simulación.

En nuestro diagrama, mostramos con línea roja continua el equilibrio entre el estado nativo y los estados desnaturalizado e hinchado. Hemos incluido además, con una línea discontinua de color negro, los puntos intermedios del tránsito entre el estado hinchado y el estado desnaturalizado, que no se incluye en los diagramas experimentales. La principal diferencia entre nuestro diagrama y los que se muestran en la Figura 5.7 se encuentra en la línea que separa el estado N y el estado D cuando $\epsilon_{ssm} \leq 0.75$. Mostramos con una línea discontinua de color rojo la forma de los diagramas de la bibliografía. La diferencia que observamos se debe al diferente comportamiento que presentan las curvas de capacidad calorífica para las presiones más bajas, puesto que para valores bajos de ϵ_{ssm} hemos notado, como comentamos al principio de este Capítulo, un comportamiento para esta proteína diferente al de las proteínas más pequeñas que hemos estudiado en capítulos anteriores (ver Figuras 3.6 y 4.2). Muy posiblemente, estas diferencias se deban, como ya hemos comentado, al mayor tamaño de esta proteína, puesto que un comportamiento parecido se apreciaba, aunque en menor medida, para la proteína 1TEN, la segunda proteína más grande de las que hemos estudiado en esta Tesis. Por

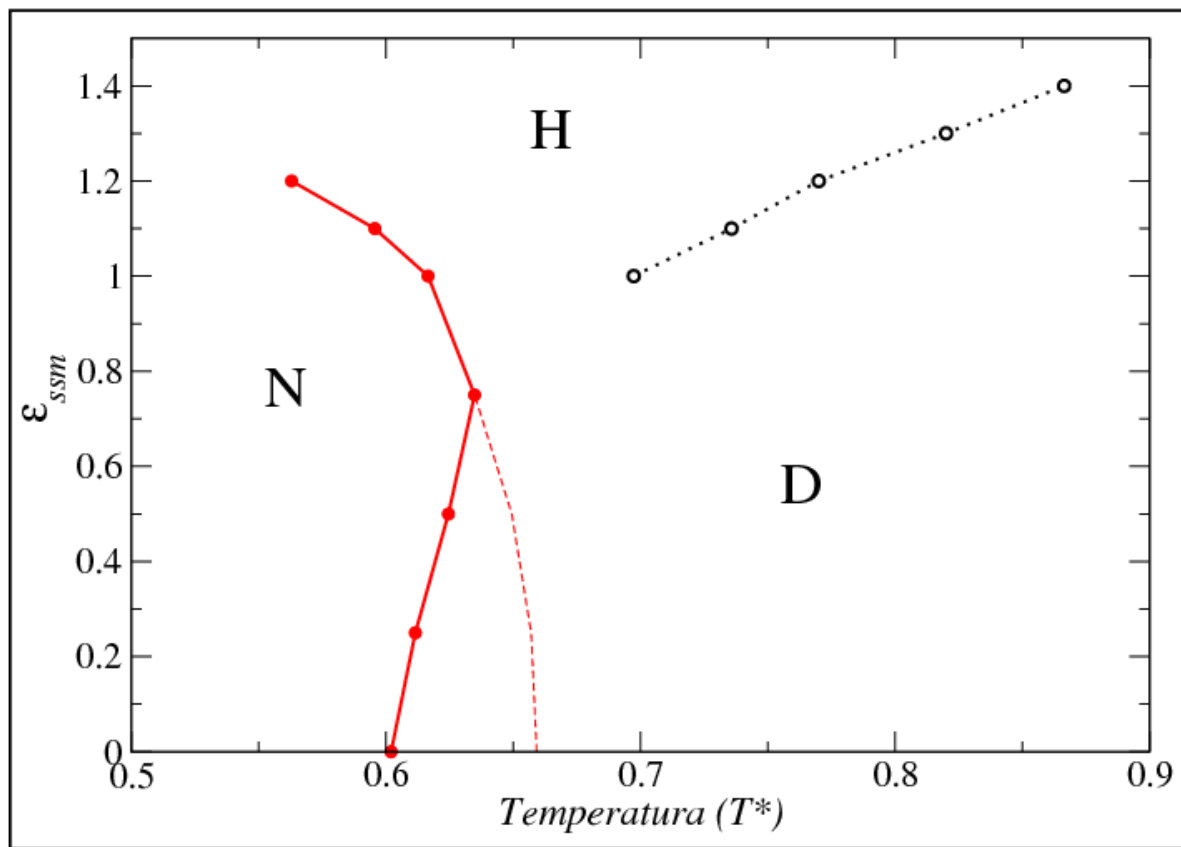


Figura 5.8: **Diagrama de fases $\epsilon_{ssm}-T$ para SNasa.** Se presenta el diagrama de fases calculado a partir de los resultados de nuestra simulación. La línea discontinua de color rojo muestra el comportamiento 'bibliográfico' para esta proteína.

tanto, parece que cuando se aplica nuestro modelo a proteínas de mayor tamaño existe un pequeño aumento de la temperatura de transición para valores bajos de ϵ_{ssm} , en comparación con el comportamiento que presentan las proteínas de menor tamaño. Como ya comentamos en la Sección 1.3.1, el tamaño de la elipse en el diagrama de fases y su cambio de orientación con respecto a los ejes de presión y temperatura es característico tanto de la proteína como de las condiciones en las que se realizan las medidas. Estas condiciones incluyen factores como el pH o la fuerza iónica, que no tienen un reflejo directo en los parámetros de nuestro modelo de simulación. Por lo tanto, es posible que la diferente orientación que muestra nuestro diagrama de fases con respecto al diagrama bibliográfico para esta proteína esté relacionado con las condiciones experimentales que se muestran en los resultados bibliográficos, que no se pueden incluir en el modelo de

simulación sencillo que hemos desarrollado.

Por último, queremos introducir un comentario sobre la zona de temperaturas bajas dentro del diagrama de fases calculado en la simulación. En nuestro diagrama no apreciamos línea de fases por encima de $\epsilon_{ssm} > 1.20$. Esto se debe a que, como ya vimos en la Figura 4.10, la línea de fases entre el estado nativo y el estado hinchado, a partir de ciertos valores de presión, se hace prácticamente paralela al eje de las abscisas. Por tanto, si el punto de corte entre el estado hinchado y el nativo cuando $\epsilon_{ssm} = 1.30$ se encuentra a una temperatura inferior a $T^* = 0.3$, el valor más pequeño al que realizamos la simulación, la temperatura de transición entre estos estados no aparece en nuestro muestreo, y por tanto no podemos seguir ampliando nuestro diagrama de fases hacia temperaturas inferiores. Por otro lado, con un modelo como el nuestro, no es posible obtener una pérdida de la estructura nativa debido a una muy baja temperatura, como sucede en la rama izquierda del diagrama experimental, en el proceso denominado *desnaturalización fría*¹⁹⁸. Este proceso está relacionado con la disminución en el orden de las moléculas de agua que rodean la proteína a temperaturas bajas, y por lo tanto es imposible reproducir este efecto en un modelo como el nuestro, que no tiene en cuenta de forma explícita la naturaleza molecular del disolvente. Por tanto, puesto que en nuestro modelo no se incluye de ninguna manera esta información, nunca podremos observar la rama izquierda del diagrama p - T , con el cierre de la elipse en el lado izquierdo del diagrama de fases.

Parece por tanto que, al menos de forma aproximada, sí que existe relación entre el diagrama de fases calculado a través de nuestra simulación y el diagrama de fases que se ha medido experimentalmente, dentro de la región en la que el modelo diseñado es capaz de reproducir el proceso de desnaturalización de una proteína.

5.6. Resumen y conclusiones del Capítulo.

A lo largo de este último Capítulo, hemos continuado validando el modelo de simulación desarrollado en esta Tesis. En este caso, hemos dado especial importancia a la comparación de los resultados obtenidos a partir de nuestra simulación con los resultados obtenidos para una misma proteína cuando se trabaja con ella experimentalmente. La proteína seleccionada ha sido SNasa, una proteína cuyo plegamiento/desnaturalización mediado por presión ha sido ampliamente estudiado por distintos grupos de investigación. Esta proteína, con estructura α - β , presenta dos características que la diferencian del resto de proteínas estudiadas en esta Tesis: su gran tamaño y que presenta un plegamiento con intermedios, es decir, se trata de una proteína que pliega en más de dos estados a presión atmosférica.

En primer lugar, hemos comprobado que nuestro modelo es capaz de reproducir este plegamiento de más de dos estados a presión atmosférica, lo cual aumenta la versatilidad del modelo que hemos desarrollado, puesto que hasta el momento todas las simulaciones de esta Tesis se habían realizado sobre proteínas que presentan plegamiento de dos estados. El poder ampliar el modelo para el estudio de procesos donde aparecen intermedios genera un valor añadido al modelo que hemos desarrollado.

Posteriormente, hemos analizado la evolución de la estabilidad del estado nativo de la proteína con el aumento de la presión. Utilizando el radio de giro como parámetro fundamental, hemos comparado nuestros estados hinchado y desnaturalizado con los dos estados desnaturalizados que se describen en la bibliografía para esta proteína, obteniendo unos resultados bastante consistentes, dentro de la comparación cualitativa de los valores del radio disponibles. Además, esta comparación nos ha permitido asignar valores de presión reales a los intervalos de nuestro parámetro ϵ_{ssm} .

Finalmente, hemos comparado nuestro diagrama de fases $\epsilon_{ssm}-T$ con el diagrama de fases $p-T$ obtenido para esta misma proteína a partir de medidas experimentales. De forma cualitativa, nuestro diagrama muestra una cierta semejanza con el diagrama

de fases medido para esta proteína, al menos para los valores más altos de presión y temperatura.

Los resultados mostrados en este Capítulo nos han servido para seguir completando los conocimientos sobre el modelo desarrollado en esta Tesis, puesto que la relación cuantitativa entre el valor que adopta ϵ_{ssm} y la presión que describe el sistema se puede trasladar a otros sistemas. Como hemos ido encontrando a lo largo de la Tesis, el valor de ϵ_{ssm} que nos permite identificar los diferentes regímenes de presión permanece prácticamente constante para los diferentes sistemas estudiados. Esto, unido al hecho de que la mayoría de las proteínas se desestabilizan dentro de unos intervalos pequeños de presión y temperatura, nos permite identificar unos intervalos de presión en función del valor que toma ϵ_{ssm} . De esta forma, conseguimos que los resultados de nuestro modelo dejen de ser exclusivamente cualitativos, y se conviertan al menos en resultados semi-cuantitativos.

El Capítulo también nos ha permitido identificar cómo cuando simulamos proteínas más grandes, estas presentan unos resultados algo diferentes a los resultados obtenidos cuando simulamos proteínas más pequeñas. Esta pequeña diferencia, que habíamos identificado para la proteína más grande simulada hasta el momento, 1TEN, se ha reproducido de forma más acusada en la proteína SNasa, proteína de mucho mayor tamaño. Es posible que la desviación observada se deba a una distinta orientación de la elipse que representa la estabilidad del estado nativo, puesto que la orientación que toma la elipse es característica de cada proteína y de las condiciones de medida, y no es factible con nuestro modelo reproducir cada una de las condiciones en las que se han realizado los diferentes experimentos. La desviación de la temperatura de transición con respecto a la tendencia observada para las proteínas de pequeño tamaño es la que nos lleva a diferencias entre el diagrama de fases $\epsilon_{ssm}-T$ que hemos calculado y el diagrama de fases $p-T$ experimental. Aun así, el diagrama de fases calculado sigue mostrando similitudes con el diagrama descrito en la bibliografía.

Concluimos, por tanto, que aunque nos hubiese gustado reproducir más exacta-

mente los resultados experimentales para esta proteína, y poder relacionar de forma cuantitativa cada una de las propiedades descritas en la bibliografía con las propiedades que describe nuestro sistema, somos conscientes de las dificultades que esto conlleva, más aún cuando trabajamos con un modelo sencillo como es el nuestro. Por tanto, la comparación semi-cuantitativa que hemos descrito a lo largo de este Capítulo nos parece un resultado satisfactorio, que nos permite la validación del modelo desarrollado.

Capítulo 6

Resumen y conclusiones generales de esta Tesis

A lo largo de esta Tesis hemos desarrollado y validado un modelo de simulación que nos permite describir, con resultados bastante satisfactorios, el proceso de desnaturalización de una proteína bajo condiciones variables de presión y temperatura. Para obtener un modelo con el que se puedan realizar cálculos de modelización molecular en tiempos razonables, hemos desarrollado un potencial de interacción de campo medio, que nos permite describir el efecto de la presión hidrostática sobre las interacciones del sistema, a pesar de que el disolvente no se considera de forma explícita en el modelo. Este potencial está basado en los denominados potenciales de tipo $G\bar{o}$, que utilizan la información presente en la geometría de la conformación nativa de la proteína, determinada experimentalmente.

En el Capítulo 1 hemos descrito la situación actual de los estudios sobre la estabilidad y desnaturalización de las proteínas por efecto de distintos agentes, poniendo especial énfasis en el efecto de la presión. Concretamente, hemos analizado las principales diferencias conocidas que existen entre los mecanismos que nos llevan hasta dos estados desnaturalizados diferentes, uno debido al efecto de la presión y otro debido al

efecto de la temperatura. Esto nos ha permitido plantear las motivaciones de nuestro trabajo y, por lo tanto, los objetivos principales de esta Tesis.

A lo largo del Capítulo 2 desarrollamos la parte metodológica más teórica de esta Tesis, definiendo tanto nuestro potencial de interacción, basado en la respuesta promedio de las interacciones hidrófobas ante un aumento de la presión hidrostática, como del modelo de simulación completo. Nuestro modelo introduce la presión de forma implícita, a través de un parámetro que hemos denominado ϵ_{ssm} , y que reproduce resultados previos sobre el efecto de la presión en las interacciones hidrófobas. Hemos adaptado esta situación a un potencial basado en la estructura nativa, con el cual se puede ya realizar la simulación de distintas proteínas a diferentes temperaturas y valores del parámetro ϵ_{ssm} . La variación de ϵ_{ssm} es la que nos permite la simulación del mismo sistema a diferentes presiones.

En los tres capítulos siguientes recogemos los resultados que hemos obtenido al aplicar el modelo desarrollado sobre 6 proteínas diferentes, permitiéndonos al mismo tiempo ir comprobando la validez de nuestro modelo. En primer lugar, en el Capítulo 3 nos hemos centrado en una única proteína para analizar en detalle la influencia de los detalles del potencial y del modelo de simulación en los resultados obtenidos. De esta forma, hemos sido capaces de encontrar una relación cualitativa entre el valor que toma el parámetro ϵ_{ssm} y la presión que adopta nuestro sistema, encontrando tres regímenes diferentes de presión: bajas, moderadas y altas, que hemos relacionado con tres intervalos del parámetro ϵ_{ssm} . Además, hemos comprobado la existencia de dos estados desnaturalizados diferentes, en función de si es la temperatura o la presión la que desnaturaliza la proteína. Hemos comprobado además que estos dos estados desnaturalizados presentan propiedades similares a las que se describen en la bibliografía para los estados desnaturalizados por presión (estado hinchado, H) y por temperatura (estado desnaturalizado térmico, D). También hemos mostrado que el modelo es capaz de reproducir diferentes tipos de tránsito entre el estado nativo y estos dos estados desnaturalizados. La posi-

bilidad de estudiar conjuntamente las variaciones de presión y temperatura durante el proceso de desnaturalización de las proteínas nos ha permitido estimar el diagrama de fases de los sistemas simulados, al menos en los intervalos de presión y temperatura accesibles a nuestra metodología, reproduciendo parte de la forma elíptica característica de los diagramas de fases p - T de las proteínas.

A continuación, a lo largo del Capítulo 4 hemos estudiado las diferencias que resultan de aplicar nuestro modelo a proteínas que representan los dos principales elementos de estructura secundaria: las hélices α y las láminas β . Hemos relacionado estas diferencias con las características específicas de los mapas de contactos nativos de cada proteína, y sobre todo con la distribución en los mismos de los enlaces de hidrógeno que estabilizan los patrones de estructura secundaria. Los resultados de este Capítulo demuestran que nuestro modelo es capaz, sin introducir ningún tipo de información añadida, de tratar de forma diferente los efectos de la presión en las hélices α y las láminas β . De esta forma, las características que adoptan dichas estructuras en el estado hinchado son diferentes entre sí, y además concuerdan bastante bien con las que se describen en la bibliografía para los dos motivos de estructura secundaria, dentro de las posibilidades del modelo.

Por último, en el Capítulo 5 realizamos una comparación más detallada entre resultados experimentales tomados de la bibliografía para una proteína dada (la nucleasa SNasa) y los resultados de nuestra simulación para la misma. A través de esta comparación de resultados, hemos comprobado que el estado hinchado que se describe experimentalmente para la proteína SNasa también es reproducido por nuestro modelo, que encuentra un estado hinchado con radio de giro superior al estado nativo e inferior al estado desnaturalizado térmico. También hemos mostrado cómo a través de los resultados que proporciona el modelo, somos capaz de reproducir parte del diagrama de fases medido para esta misma proteína. Estos resultados sirven para concluir la validación del modelo que hemos desarrollado, y además nos permite establecer una relación

cuantitativa entre el valor del parámetro ϵ_{ssm} y la presión experimental.

En resumen, a lo largo del trabajo recogido en esta Tesis hemos mostrado cómo un modelo sencillo de modelización molecular es capaz de reproducir las características principales de la desnaturalización de una proteína por el efecto de un aumento de la presión hidrostática en el sistema. Hemos demostrado que el modelo reproduce la aparición de un estado desnaturalizado hinchado, con características diferentes a las del estado desnaturalizado por efecto de la temperatura. De esta forma, hemos validado tanto las características del modelo como la metodología de simulación desarrollados en este trabajo, permitiendo su uso futuro para el análisis de otros sistemas de relevancia.

Capítulo 7

Summary

“Coarse-grained simple models to study the effect of pressure in protein folding processes”

Author: Ramiro Perezzan Rodriguez.

Supervisor: Antonio Rey Gayo.

7.1. Introduction.

From the beginning of the 20th century, the effect of pressure on the native conformation of proteins is known. A hydrostatic pressure of about 3 kbar can unfold a protein³⁵, and therefore it was thought that high pressures can be somehow similar to the effect of temperature on the stability of these biopolymers³⁵. Nevertheless, the difficult task to carry out accurate experimental measurements at high pressures, especially in a biological environment, has made the pressure effects to be traditionally less known and understood than other properties which also denature natural proteins, as temperature, pH or chemical agents. In the last decade, however, the continuous development of biophysical techniques has led several research groups to pursue the

study of pressure effects on proteins, from different perspectives. On one hand, there is a fundamental need to understand the interactions responsible for the stability of folded proteins^{38,39}, and the pressure represents in this sense a novel tool, which provides new insights into the study of the folding funnel. On the other hand, there are also several practical applications of pressure in the fields of pharmacology⁷¹, food industry^{74,75}, or even medicine^{64,66}, since high pressures can inhibit protein aggregation⁶³ related to the development of neurodegenerative diseases such as Alzheimer or Parkinson^{36,68}. It is very important to mention that, although the unfolded state is sometimes considered as “everything which is not native”, both experimental and numerical simulation studies have shown that there are differences between protein unfolded states resulting from a pressure increase or from a temperature increase^{39,49}. For single domain proteins, the temperature unfolding process is considered to be cooperative for many proteins, and both the tertiary and the secondary structures of the native conformation are essentially lost in the unfolded state. On the other hand, a pressure unfolded protein shows a larger radius of gyration than the native state, but the secondary structure, or at least a substantial fraction of it, is preserved. Therefore, this unfolded state keeps a certain globular shape, but it is swollen in comparison to the native conformation^{49,50}. Computational models which use molecular simulations to analyze the effect of pressure on protein stability have mostly used detailed, all-atom models with explicit solvent and molecular dynamics techniques^{95–97}. These models, as accurate as they are, may be also too detailed and computationally costly to properly consider (i.e., with enough statistical accuracy) processes that happen in long time scales, as the full folding of a protein. Coarse-grained simple models, on the other hand, allow exploring the folding/unfolding equilibrium of a protein in a reasonable computational time^{114,131}. The loss of the atomic details in the simulation results becomes, in this sense, an affordable compromise, especially when the simple model is well defined for the specific purpose of a given investigation. Although the pressure unfolding of proteins seems to be incompatible with

molecular models where the packing details have to be carefully considered, it is still possible to obtain interesting physical insights of this process by using reduced representations of the system and mean field potentials¹¹⁹. They have been commonly used, with different levels of approach, to describe the stabilizing interactions responsible for a protein native conformation, mainly the hydrophobic interactions and their relation to the solvation of the hydrophobic residues, with the solvent being implicitly considered^{112,113}. Therefore, it seems natural as a first approach, at the level of a coarse-grained model, to simulate the effect of pressure by a change in the different contributions to the potential of mean force.

7.2. Main Objectives.

In this Thesis, we have tried to design a coarse grained model which uses a carefully defined structure-based potential to analyze the effects of hydrostatic pressure on the unfolding transition of globular proteins. We want to include this model in a Monte Carlo simulation algorithm, based on the previous experience of our research group^{102,148}, to study the stability of the protein native state under increases in both temperature and pressure. Our aim is, therefore, to try to find the pressure-temperature phase diagram of the simulated proteins^{40,42}, analyzing the states involved in its elliptical shape, to check if we are able to reproduce the experimentally found behavior. Our simulation model has also to allow us analyzing the structural characteristics of the pressure-induced unfolded state. The comparison of all these simulation results with the experimental evidence available from the literature will be the final test of the designed model.

7.3. Methods.

The interaction potential we have designed in this Doctoral Thesis results from the potential of mean force between two hydrophobic molecules, obtained when the degrees

of freedom of the surrounding water are integrated out. It shows two attractive basins: one, at short distances, corresponds to the direct contact between the interacting units; another one, at a larger distance, represents the interaction between the same units mediated by a single water molecule, and is named “solvent separated minimum”, ssm, along this work^{122,123}. These two minima appear at a distance of 3 Å, which roughly corresponds to the diameter of a water molecule. The minima are separated by a region of positive potential energy, named desolvation barrier, which represents the energy needed to expel the last water molecule in the approaching of the considered hydrophobic units. Essentially, the pressure induces changes in the relative depth of the minima and the height of the barrier. To favor a systematic analysis of this model, we have used the depth of the direct contact as our energy unit, and carried out an increase in the depth of the ssm minimum, ϵ_{ssm} , to reproduce the effect of an increase in the hydrostatic pressure of the system??.

These ideas are introduced into a structure-based model, whose stabilizing interactions (among residues which form a contact in the native structure) are made pressure-dependent through the strategy just described. In the initial stages of the model development, we cannot establish a real pressure scale, but we will be able to gradually increase both pressure and temperature to analyze the characteristics of the resulting folding/unfolding transition. Since in this work we are interested in its thermodynamic and structural features, we have used a Monte Carlo sampling procedure¹⁰³, which allows us a simple mathematical definition of the interaction potential described (we do not need the potential to be differentiable, as it happens in molecular or stochastic dynamics simulations). The model, as required, is numerically efficient, and has allowed us to analyze in detail the influence of pressure on the stability of the native state, and to find pressure unfolded states different from those resulting from thermal denaturation.

7.4. Results.

First place, we have carried out a systematic tuning of the interaction potential and simulation model by analyzing the folding/unfolding transition of protein 2GB1¹⁴⁹ (a name taken from its PDB¹³⁷ code, which serves as the basis to define the contact map of the native interactions). We have computed the heat capacity curves as a function of temperature for the different values of ϵ_{ssm} simulated, each one corresponding to a different pressure. An increase in pressure leads to a destabilization of the folded structure, as shown by the shift of the transition temperature towards lower values. The characteristics of the transition are also affected by pressure. At low pressures, the typical narrow heat capacity peak corresponding to a cooperative transition is found²⁶. However, at higher pressures a wider peak is observed. A further analysis of the transition through the calculation of the profiles for the Helmholtz energy clearly shows that the barrier between the folded and the unfolded states becomes lower when the pressure increases, resulting in a less cooperative transition³⁰. Moreover, the unfolded state clearly changes from a high energy denatured state at low pressures, with a broad set of mostly unstructured conformations, to a state with more contacts, as shown by its lower energy. An analysis of the geometrical features of this unfolded state appearing at high pressures shows that it is relatively compact, as measured by its average radius of gyration, and not so different from the native state, as measured by its average RMSD value (“root mean square deviation”, a metrics of the dissimilarity between the experimental native structure and any of the simulated conformations). We have therefore defined in our simulations a “swollen” state, as the state resulting from the denaturation at high pressures and room temperature, which is clearly different from the native and the thermally denatured states. The possibility to find well defined ranges for the aforementioned structural properties has allowed us to quantitatively evaluate the populations of the three different states which appear along our sampling, depending on the simulated conditions. This, in turn, permits us to determine the conditions which correspond to

equivalent populations between pairs of states. This way, we have been able to establish a set of points which characterize the transition between the native and the thermally unfolded state, between the native and the swollen state, and between the swollen and the thermally unfolded state. The representation of these points on a graph which uses temperature and ϵ_{ssm} as its axes represents our sketch of the pressure-temperature phase diagram for the simulated protein.

We have extended our simulations to four additional proteins, two with an all- α native structure and another two with an all- β native fold. In all the cases, we have obtained a similar set of folding/unfolding transitions, and also equivalent phase diagrams. In this part of the analysis, we have tried to get a deeper insight into the structural characteristics of the swollen states for the different proteins^{93,166}. It is important to recall that we have not changed anything in the model or the simulation conditions from one protein to another. Therefore, the differences in behavior we have found are somehow encoded into the contact map of the corresponding native structures. We have found that the α structures are less sensitive to pressure than β ones. This fact is related to the possibility in our model to preserve helical fragments up to relatively high pressures, since hydrogen bond contacts which stabilize these secondary structure elements do not have the physical possibility to inject a molecule of water between the interacting residues. The native contacts between different helices, on the other hand, switch from the direct contact minimum of our potential to the solvent separated one, this way allowing for a swollen state which clearly resembles an expanded version of the native conformation. In β structures, the different types of native contacts are much more mixed up in the contact maps, something what makes these structures more prone to pressure denaturation, since the possible swollen states do not easily reflect a well-defined β -sheet.

Finally, we have directly compared the results of our simulations to experimental results, to try to better check the validity of our model. It is no easy to find experimental

results which we can compare, on a quantitative basis, to our simulation data. We have finally decided to use a nuclease from staphylococcus, SNase, a protein which has been extensively analyzed under pressure^{189,192,194}. A few of the results published for this protein can be directly related to the structural properties we compute in our simulations^{45,188}. We have considered the radius of gyration, one of the properties we had already used in the analysis of our simulations. Although the values of these radii experimentally reported in the literature are, as a matter of fact, hydrodynamic radii which cannot be quantitative compared to our real radii of gyration¹⁹⁵, the changes in this property as a function of pressure for a given temperature have allowed us to establish a relationship between the experimental values of pressure and our scale of ϵ_{ssm} . Moreover, this relationship found for SNase seems to be general for different proteins, at least for those analyzed in this Thesis, and can be therefore a general feature of our model. The calculation of the phase diagram for this protein has shown features than, in our model, seem to be slightly different between small and medium-size proteins. In the latter case, a moderate pressure implies a certain stabilization of the native state. Although the experimental phase diagrams we have found for SNase do not show this behavior⁴¹, it has been indeed reported in the literature for different proteins, and could be related to different experimental conditions (pH, ionic strength, etc.) which affect the orientation of the elliptic diagram with respect to the pressure-temperature axes^{42,43}.

7.5. Conclusions.

In this work, we have shown the performance of a very simple simulation model we have introduced in this Thesis to study the pressure unfolding of proteins at a coarse-grained level. By a thorough analysis of the simulation results, we have checked that it is indeed possible to reproduce the unfolding behavior of a protein by the effect of either temperature or pressure using the same model. Initially, this latter property is defined at the level of a model parameter, although the comparison to experimental

results allows us establishing a quantitative relationship to real pressures. Moreover, the model reproduces a pressure-induced unfolded state, which we have named “swollen” state, with structural characteristics which make it rather different from the thermally denatured state. The model even shows distinctive features for this swollen state for pressure unfolded all- α and all- β proteins, a remarkable result having into account the simplicity of the model and the identical treatment we use for both types of structural families. Probably the most impressive result for the model is the possibility to reproduce part of the pressure-temperature phase diagram of a globular protein, clearly showing a fraction of the characteristic ellipsis which has been experimentally found^{??}. Therefore, although the model shows limitations related to its simplicity, which have discussed along this Thesis, we consider we have fully validated its capabilities, as it was our main objective at the beginning of this work. The model designed can therefore be applied to analyze, at a modest computational cost, the pressure folding/unfolding transition of globular proteins.

Bibliografía

1. Curtis, H., Barnes, N. S., Schneek, A. & Flores, G., *Biología* (Editorial Médica Panamericana, Buenos Aires, 2000).
2. Finkelstein, A. V. & Ptitsyn, O. B., *Protein physics. A course of lectures* (Academic Press, Amsterdam, 2002).
3. Soto Camara, J. L., *Química orgánica. 1, Conceptos básicos* (Síntesis, Madrid, 2003).
4. Mathews, C. K., Holde, V., Ahern, K. G. & Gonzalez de Buitrago, J. M., *Bioquímica* (Addison Wesley, Madrid, 2002).
5. Dill, K. A. & Bromberg, S., *Molecular driving forces* (Garland Science, New York, 2010).
6. Nelson, D. L., Cox, M. M., Cuchillo, C. & Lehninger, A. L., *Lehninger Principios de bioquímica* (Omega, Barcelona).
7. Gómez-Moreno Calera, C. & Sancho Sanz, J., *Estructura de proteínas* (Ariel, Barcelona, 2003).
8. Yasuda, S., Yoshidome, T., Oshima, H., Kodama, R., Harano, Y. & Kinoshita, M., Effects of side-chain packing on the formation of secondary structures in protein folding. *J. Chem. Phys.* **132**, 65105 (2010).

9. Anfinsen, C. B., Principles that govern the folding of protein chains. *Science* **181**, 223 (1973).
10. Anfinsen, C. B. & Scheraga, H. A., Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.* **29**, 205 (1975).
11. Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R., The protein folding problem. *Annu. Rev. Biophys.* **37**, 289 (2008).
12. Levinthal, C., Are there pathways for protein folding. *J. Chim. Phys.* **65**, 44 (1968).
13. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G., The energy landscapes and motions of proteins. *Science* **254**, 1598 (1991).
14. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G., Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167 (1995).
15. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G., Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545 (1997).
16. Jahn, T. R. & Radford, S. E., The yin and yang of protein folding. *FEBS J.* **272**, 5962 (2005).
17. Onuchic, J. N. & Wolynes, P. G., Theory of protein folding. *Curr. Opin. Struct. Biol.* **14**, 70 (2004).
18. Bryngelson, J. D. & Wolynes, P. G., Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7524 (1987).
19. Brooks III, C. L., Gruebele, M., Onuchic, J. N. & Wolynes, P. G., Chemical physics of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11037 (1998).
20. Wolynes, P. G., Eaton, W. A. & Fersht, A. R., Chemical physics of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17770 (2012).

21. Clementi, C., Nymeyer, H. & Onuchic, J. N., Topological and energetic factors: what determines the structural details of the transition state ensemble and 'En-route' intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937 (2000).
22. Dill, K. A., Dominant forces in protein folding. *Biochemistry* **29**, 7133 (1990).
23. Nicholls, A., Sharp, K. A. & Honig, B., Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281 (1991).
24. Dill, K. A., Polymer principles and protein folding. *Protein Sci.* **8**, 1166 (1999).
25. Cordes, M. H., Davidson, A. R. & Sauer, R. T., Sequence space, folding and protein design. *Curr. Opin. Struct. Biol.* **6**, 3 (1996).
26. Jackson, S. E., How do small single-domain proteins fold? *Fold. Des.* **3**, 81 (1998).
27. Cecconi, C., Shank, E. A., Bustamante, C. & Marqusee, S., Direct observation of the three-state folding of a single protein molecule. *Science* **309**, 2057 (2005).
28. Privalov, P. L., Intermediate states in protein folding. *J. Mol. Biol.* **258**, 707 (1996).
29. Garcia-Mira, M. M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J. M. & Muñoz, V., Experimental identification of downhill protein folding. *Science* **298**, 2191 (2002).
30. Naganathan, A. N., Perez-Jimenez, R., Sanchez-Ruiz, J. M. & Muñoz, V., Robustness of downhill folding: guidelines for the analysis of equilibrium folding experiments on small proteins. *Biochemistry* **44**, 7435 (2005).
31. Lilley, D. M. J., Clore, M., Campbell, S. & Hubbard, R. E., *Protein Folding, Misfolding and Aggregation*, RSC Biomolecular Sciences (The Royal Society of Chemistry, London, 2008).

32. Giancoli, D. C., *Physics: principles with applications* (Pearson, New Jersey, 2005).
33. Huston, R. L., Josephs & Blake, A., *Practical stress analysis in engineering design*. (CRC, Boca Raton, 2008).
34. Menéndez, J. M., *Materia a alta presion: fundamentos y aplicaciones* (Servicio de Publicaciones, Universidad de Oviedo, 2011).
35. Bridgman, P., The coagulation of albumen by pressure. *J. Biol. Chem.* **19**, 511 (1914).
36. Meersman, F., Dobson, C. M. & Heremans, K., Protein unfolding, amyloid fibril formation and configurational energy landscapes under high pressure conditions. *Chem. Soc. Rev.* **35**, 908 (2006).
37. Boldyreva, E. & Dera, P., *High-pressure crystallography: from fundamental phenomena to technological applications* (Springer, Dordrecht, 2010).
38. Rick, S., Free energy, entropy and heat capacity of the hydrophobic interaction as a function of pressure. *J. Chem. Phys. B* **104**, 6884 (2000).
39. Harano, Y., Yoshidome, T. & Kinoshita, M., Molecular mechanism of pressure denaturation of proteins. *J. Chem. Phys.* **129**, 145103 (2008).
40. Smeller, L., Pressure-temperature phase diagrams of biomolecules. *Biochim. Biophys. Acta* **1595**, 11 (2002).
41. Panick, G., Vidugiris, G. J., Malessa, R., Rapp, G., Winter, R. & Royer, C. A., Exploring the temperature-pressure phase diagram of staphylococcal nuclease. *Biochemistry* **38**, 4157 (1999).
42. Meersman, F., Smeller, L. & Heremans, K., Protein stability and dynamics in the pressure-temperature plane. *Biochim. Biophys. Acta* **1764**, 346 (2006).

43. Winter, R., Lopes, D., Grudzielanek, S. & Vogtt, K., Towards an understanding of the temperature/pressure configurational and free-energy landscape of biomolecules. *J. Non-Equilib. Thermodyn.* **32**, 41 (2007).
44. Maeno, A., Matsuo, H. & Akasaka, K., The pressure-temperature phase diagram of hen lysozyme at low pH. *Biophysics* **5**, 1 (2009).
45. Panick, G., Malessa, R., Winter, R., Rapp, G., Frye, K. J. & Royer, C. A., Structural characterization of the pressure-denatured state and unfolding/refolding kinetics of staphylococcal nuclease by synchrotron small-angle X-ray scattering and Fourier-transform infrared spectroscopy. *J. Mol. Biol.* **275**, 389 (1998).
46. Dill, K. A. & Shortle, D., Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795 (1991).
47. Ptitsyn, O. B., Molten globule and protein folding. *Adv. Protein Chem.* **47**, 83 (1995).
48. Yoshidome, T., Harano, Y. & Kinoshita, M., Pressure effects on structures formed by entropically driven self-assembly: illustration for denaturation of proteins. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **79**, 11912 (2009).
49. Sarupria, S., Ghosh, T., García, A. E. & Garde, S., Studying pressure denaturation of a protein by molecular dynamics simulations. *Proteins* **78**, 1641 (2010).
50. Yoshidome, T., General framework of pressure effects on structures formed by entropically driven self-assembly. *Entropy* **12**, 1632 (2010).
51. Crothers, D. M., Cole, P. E., Hilbers, C. W. & Shulman, R. G., The molecular mechanism of thermal unfolding of Escherichia coli formylmethionine transfer RNA. *J. Mol. Biol.* **87**, 63 (1974).

52. Baldwin, R. L., Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8069 (1986).
53. Boonyaratanakornkit, B. B., Park, C. B. & Clark, D. S., Pressure effects on intra- and intermolecular interactions within proteins. *Biochim. Biophys. Acta* **1595**, 235 (2002).
54. Royer, C., A discussion of the physical basis for the pressure unfolding of proteins. *Trends in high pressure bioscience and biotechnology, proceedings* **19**, 17 (2002).
55. Royer, C. A., Revisiting volume changes in pressure-induced protein unfolding. *Biochim. Biophys. Acta* **1595**, 201 (2002).
56. Rouget, J. B., Aksel, T., Roche, J., Saldana, J.-L., Garcia, A. E., Barrick, D. & Royer, C. A., Size and sequence and the volume change of protein folding. *J. Am. Chem. Soc.* **133**, 6020 (2011).
57. Chalikian, T. V. & Macgregor, R. B., Origins of pressure-induced protein transitions. *J. Mol. Biol.* **394**, 834 (2009).
58. Roche, J., Caro, J. A., Norberto, D. R., Barthe, P., Roumestand, C., Schlessman, J. L., Garcia, A. E., García-Moreno, B. E. & Royer, C. A., Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6945 (2012).
59. Grigera, J. R. & McCarthy, A. N., The behavior of the hydrophobic effect under pressure and protein denaturation. *Biophys. J.* **98**, 1626 (2010).
60. Kinoshita, M., Importance of translational entropy of water in biological self-assembly processes like protein folding. *Int. J. Mol. Sci.* **10**, 1064 (2009).
61. Frye, K. J. & Royer, C. A., Probing the contribution of internal cavities to the volume change of protein unfolding under pressure. *Protein Sci.* **7**, 2217 (1998).

-
62. Kamatari, Y. O., Smith, L. J., Dobson, C. M. & Akasaka, K., Cavity hydration as a gateway to unfolding: an NMR study of hen lysozyme at high pressure and low temperature. *Biophys. Chem.* **156**, 24 (2011).
63. Foguel, D., Robinson, C. R., de Sousa, P. C., Silva, J. L. & Robinson, A. S., Hydrostatic pressure rescues native protein from aggregates. *Biotechnol. Bioeng.* **63**, 552 (1999).
64. Meersman, F. & Dobson, C. M., Probing the pressure-temperature stability of amyloid fibrils provides new insights into their molecular properties. *Biochim. Biophys. Acta* **1764**, 452 (2006).
65. Dobson, C. M., The structural basis of protein folding and its links with human disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 133 (2001).
66. Chiti, F. & Dobson, C. M., Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333 (2006).
67. Dobson, C. M., Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329 (1999).
68. Dobson, C. M., Protein folding and misfolding. *Nature* **426**, 884 (2003).
69. Foguel, D., Suarez, M. C., Ferrão-Gonzales, A. D., Porto, T. C. R., Palmieri, L., Einsiedler, C. M., Andrade, L. R., Lashuel, H. A., Lansbury, P. T., Kelly, J. W. & Silva, J. L., Dissociation of amyloid fibrils of alpha-synuclein and transthyretin by pressure reveals their reversible nature and the formation of water-excluded cavities. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9831 (2003).
70. Masson, P., Tonello, C. & Balny, C., High-pressure biotechnology in medicine and pharmaceutical science. *J. Biomed. Biotechnol.* **1**, 85 (2001).

71. Silva, J., Foguel, D., Suarez, M., Gomes, A. & Oliveira, A., High-pressure applications in medicine and pharmacology. *J. Phys. Condens. Matter* **16**, 929 (2004).
72. Cohen, P., Protein kinases—the major drug targets of the twenty-first century? *Nat. Rev. Drug. Discov.* **1**, 309 (2002).
73. Jameel, F. & Hershenson, S., *Formulation and process development strategies for manufacturing biopharmaceuticals* (Wiley, Hoboken, 2010).
74. Knorr, D., Heinz, V. & Buckow, R., High pressure applications for food biopolymers. *Biochim. Biophys. Acta* **1764**, 619 (2006).
75. Liu, X., Ning, J. & Clark, S., Changes in structure and functional properties of whey proteins induced by high hydrostatic pressure: A review. *Front. Chem. Engin. China* **3**, 436 (2009).
76. Cheftel, J. C. & Culioli, J., Effects of high pressure on meat: A review. *Meat Sci.* **46**, 211 (1997).
77. Hillson, N., Onuchic, J. N. & García, A. E., Pressure-induced protein-folding/unfolding kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14848 (1999).
78. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M. V., Alonso, D. O. V., Daggett, V. & Fersht, A. R., The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421**, 863 (2003).
79. Mehta, A., Rief, M., Spudich, J., Smith, D. & Simmons, R., Single-molecule biomechanics with optical methods. *Science* **283**, 1689 (1999).
80. Schuler, B., Lipman, E. A. & Eaton, W. A., Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743 (2002).

81. Muir, T. W., Sondhi, D. & Cole, P. A., Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6705 (1998).
82. *Protein folding handbook Vol. 1 Vol. 5* (Wiley-VCH, Weinheim, 2005).
83. McGuffin, L. J., Bryson, K. & Jones, D. T., The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404 (2000).
84. Moult, J., A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* **15**, 285 (2005).
85. Friesner, R. A., Prigogine, I. & Rice, S. A., *Computational methods for protein folding* (Wiley, New York, 2002).
86. Collins, M. D., Kim, C. U. & Gruner, S. M., High-pressure protein crystallography and NMR to explore protein conformations. *Annu Rev Biophys* **40**, 81 (2011).
87. Winter, R., Synchrotron x-ray and neutron small-angle scattering of lyotropic lipid mesophases, model biomembranes and proteins in solution at high pressure. *Biochim. Biophys. Acta* **1595**, 160 (2002).
88. Kim, C. U., Hao, Q. & Gruner, S. M., Solution of protein crystallographic structures by high-pressure cryocooling and noble-gas phasing. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 687 (2006).
89. Kim, C. U., Wierman, J. L., Gillilan, R., Lima, E. & Gruner, S. M., A high-pressure cryocooling method for protein crystals and biological samples with reduced background x-ray scatter. *J. Appl. Crystallogr.* **46**, 234 (2013).
90. Higashiura, A., Ohta, K., Masaki, M., Sato, M., Inaka, K., Tanaka, H. & Nakagawa, A., High-resolution X-ray crystal structure of bovine H-protein using the high-pressure cryocooling method. *J. Synchrotron. Radiat.* **20**, 989 (2013).

91. Jonas, J. & Jonas, A., High-pressure NMR spectroscopy of proteins and membranes. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 287 (1994).
92. Inoue, K., Yamada, H., Imoto, T. & Akasaka, K., High pressure NMR study of a small protein, gurmarin. *J. Biomol. NMR* **12**, 535 (1998).
93. Wilton, D. J., Tunncliffe, R. B., Kamatari, Y. O., Akasaka, K. & Williamson, M. P., Pressure-induced changes in the solution structure of the GB1 domain of protein G. *Proteins* **71**, 1432 (2008).
94. Kamatari, Y. O., Kitahara, R., Yamada, H., Yokoyama, S. & Akasaka, K., High-pressure NMR spectroscopy for characterizing folding intermediates and denatured states of proteins. *Methods* **34**, 133 (2004).
95. Kitchen, D. B., Reed, L. H. & Levy, R. M., Molecular dynamics simulation of solvated protein at high pressure. *Biochemistry* **31**, 10083 (1992).
96. Paschek, D., Gnanakaran, S. & Garcia, A. E., Simulations of the pressure and temperature unfolding of an alpha-helical peptide. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6765 (2005).
97. Okumura, H., Temperature and pressure denaturation of chignolin: folding and unfolding simulation by multibaric-multithermal molecular dynamics method. *Proteins* **80**, 2397 (2012).
98. Paci, E. & Marchi, M., Intrinsic compressibility and volume compression in solvated proteins by molecular dynamics simulation at high pressure. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11609 (1996).
99. Collins, M. D., Hummer, G., Quillin, M. L., Matthews, B. W. & Gruner, S. M., Cooperative water filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16668 (2005).

100. Cheung, M. S., García, A. E. & Onuchic, J. N., Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 685 (2002).
101. Shea, J. E. & Brooks III, C. L., From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **52**, 499 (2001).
102. Prieto, L., *Potenciales basados en topología para el estudio de las transiciones de plegamiento y agregación de proteínas*, Tesis doctoral, Universidad Complutense de Madrid, Fac. CC. Químicas, Madrid (2008).
103. Frenkel, D. & Smit, B., *Understanding molecular simulation from algorithms to applications* (Academic Press, San Diego, 2002).
104. Schlick, T., *Molecular modeling and simulation. An interdisciplinary guide* (Springer Media, New York, 2010).
105. Zwier, M. C. & Chong, L. T., Reaching biological timescales with all-atom molecular dynamics simulations. *Curr. Opin. Pharmacol.* **10**, 745 (2010).
106. Kubelka, J., Hofrichter, J. & Eaton, W. A., The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **14**, 76 (2004).
107. Pande, V., Sorin, E., Snow, C. & Rhee, Y., Computer simulations of protein folding, in *Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches*, 85 (The Royal Society of Chemistry, London, 2008).
108. Ponder, J. W. & Case, D. A., Force fields for protein simulations. *Adv. Protein Chem.* **66**, 27 (2003).
109. Tozzini, V., Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **15**, 144 (2005).

-
110. Noid, W. G., Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **139**, 90901 (2013).
111. http://www.sebbm.es/archivos_tinymce/julio2011_lidiaprieto.pdf .
112. Taketomi, H., Ueda, Y. & Gō, N., Studies on protein folding, unfolding and fluctuations by computer simulation. i. the effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* **7**, 445 (1975).
113. Gō, N., Protein folding as a stochastic-process. *J. Stat. Phys.* **30**, 413 (1983).
114. Prieto, L., de Sancho, D. & Rey, A., Thermodynamics of Gō-type models for protein folding. *J. Chem. Phys.* **123**, 154903 (2005).
115. Larriva, M., Prieto, L., Bruscolini, P. & Rey, A., A simple simulation model can reproduce the thermodynamic folding intermediate of apoflavodoxin. *Proteins* **78**, 73 (2010).
116. Hills, R. D. & Brooks III, C. L., Insights from coarse-grained Gō models for protein folding and dynamics. *Int. J. Mol. Sci.* **10**, 889 (2009).
117. Prieto, L. & Rey, A., Influence of the native topology on the folding barrier for small proteins. *J. Chem. Phys.* **127**, 175101 (2007).
118. Prieto, L. & Rey, A., Influence of the chain stiffness on the thermodynamics of a Gō type model for protein folding. *J. Chem. Phys.* 165103 (2007).
119. Roux, B., The calculation of the potential of mean force using computer-simulations. *Comput. Phys. Comm.* **91**, 275 (1995).
120. Roux, B. & Simonson, T., Implicit solvent models. *Biophys. Chem.* **78**, 1 (1999).
121. Tanford, C., How protein chemists learned about the hydrophobic factor. *Protein Sci.* **6**, 1358 (1997).

122. Czaplewski, C., Rodziewicz-Motowidło, S., Liwo, A., Ripoll, D. R., Wawak, R. J. & Scheraga, H. A., Molecular simulation study of cooperativity in hydrophobic association. *Protein Sci.* **9**, 1235 (2000).
123. Moghaddam, M. S., Shimizu, S. & Chan, H. S., Temperature dependence of three-body hydrophobic interactions: potential of mean force, enthalpy, entropy, heat capacity, and nonadditivity. *J. Am. Chem. Soc.* **127**, 303 (2005).
124. Liu, Z. & Chan, H. S., Solvation and desolvation effects in protein folding: native flexibility, kinetic cooperativity and enthalpic barriers under isostability conditions. *Phys. Biol.* **2**, S75 (2005).
125. Liu, Z. & Chan, H. S., Desolvation is a likely origin of robust enthalpic barriers to protein folding. *J. Mol. Biol.* **349**, 872 (2005).
126. Ghosh, T., García, A. E. & Garde, S., Molecular dynamics simulations of pressure effects on hydrophobic interactions. *J. Am. Chem. Soc.* **123**, 10997 (2001).
127. Hummer, G., Garde, S., García, A. E., Paulaitis, M. E. & Pratt, L. R., The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1552 (1998).
128. Ghosh, T., Garcia, A. & Garde, S., Enthalpy and entropy contributions to the pressure dependence of hydrophobic interactions. *J. Chem. Phys.* **116**, 2480 (2002).
129. Lammert, H., Schug, A. & Onuchic, J. N., Robustness and generalization of structure-based models for protein folding and function. *Proteins* **77**, 881 (2009).
130. Noel, J. K., Whitford, P. C. & Onuchic, J. N., The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B* **116**, 8692 (2012).

-
131. Enciso, M. & Rey, A., A refined hydrogen bond potential for flexible protein models. *J. Chem. Phys.* **132**, 235102 (2010).
132. Enciso, M. & Rey, A., Improvement of structure-based potentials for protein folding by native and nonnative hydrogen bonds. *Biophys. J.* **101**, 1474 (2011).
133. DeLano, W. & Lam, J., PyMOL: A communications tool for computational models. *Abs. Pap.: Am. Chem. Soc.* **230**, 1371 (2005).
134. Hansmann, U., Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **281**, 140 (1997).
135. Sugita, Y. & Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141 (1999).
136. Allen, M. P. & Tildesley, D. J., *Computer simulation of liquids* (Clarendon Press, Oxford, 1989).
137. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D. & Zardecki, C., The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 899 (2002).
138. Gront, D., Kolinski, A. & Skolnick, J., Comparison of three Monte Carlo conformational search strategies for a protein-like homopolymer model: Folding thermodynamics and identification of low-energy structures. *J. Chem. Phys.* **113**, 5065 (2000).
139. Humphrey, W., Dalke, A. & Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33 (1996).
140. <http://www.nongnu.org/xmakemol> .

-
141. Grosberg, A. I. & Khokhlov, A. R., *Statistical physics of macromolecules* (AIP Press, New York, 1994).
142. Fixman, M., Radius of gyration of polymer chains. *J. Chem. Phys.* **36**, 306 (1962).
143. Cohen, F. E. & Sternberg, M. J., On the prediction of protein structure: The significance of the root-mean-square deviation. *J. Mol. Biol.* **138**, 321 (1980).
144. Brüschweiler, R., Efficient RMSD measures for the comparison of two molecular ensembles. root-mean-square deviation. *Proteins* **50**, 26 (2003).
145. Ibarra-Molero, B. & Sanchez-Ruiz, J. M., Statistical differential scanning calorimetry: Probing protein folding-unfolding ensembles, in *Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches*, 85 (The Royal Society of Chemistry, London, 2008).
146. Kumar, S., Bouzida, D., Swendsen, R., Kollman, P. & Rosenberg, J., The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comput. Chem* **13**, 1011 (1992).
147. Chodera, J. D., Swope, W. C., Pitera, J. W., Seok, C. & Dill, K. A., Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J. Chem. Theory Comp.* **3**, 26 (2007).
148. Larriva, M., *Estructura nativa y secuencia de el proceso de plegamiento de proteínas globulares*, Tesis doctoral, Universidad Complutense de Madrid, Fac. CC. Químicas, Madrid (2010).
149. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T. & Clore, G. M., A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **253**, 657 (1991).

-
150. Ding, K., Louis, J. M. & Gronenborn, A. M., Insights into conformation and dynamics of protein GB1 during folding and unfolding by NMR. *J. Mol. Biol.* **335**, 1299 (2004).
151. Best, R. B. & Mittal, J., Free-energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: Similarities and differences. *Proteins* **79**, 1318 (2011).
152. Shimada, J. & Shakhnovich, E., The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11175 (2002).
153. McCallister, E. L., Alm, E. & Baker, D., Critical role of beta-hairpin formation in protein G folding. *Nat. Struct. Biol.* **7**, 669 (2000).
154. Derreumaux, P., Role of supersecondary structural elements in protein G folding. *J. Chem. Phys.* **119**, 4940 (2003).
155. Prieto, L. & Rey, A., Topology-based potentials and the study of the competition between protein folding and aggregation. *J. Chem. Phys.* **130**, 115101 (2009).
156. Reimer, U. & Fischer, G., Local structural changes caused by peptidyl-prolyl cis/trans isomerization in the native state of proteins. *Biophys. Chem.* **96**, 203 (2002).
157. Vonhippel, P. & Wong, K., Neutral salts - Generality of their effects on stability of macromolecular conformations. *Science* **145**, 577 (1964).
158. Wedemeyer, W. J., Welker, E., Narayan, M. & Scheraga, H. A., Disulfide bonds and protein folding. *Biochemistry* **39**, 4207 (2000).
159. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. & Thornton, J. M., CATH—a hierarchic classification of protein domain structures. *Structure* **5**, 1093 (1997).

-
160. Payne, V., Matubayasi, N., Murphy, L. & Levy, R., Monte Carlo study of the effect of pressure on hydrophobic association. *J. Chem. Phys. B* **101**, 2054 (1997).
161. Gómez, J., Hilser, V. J., Xie, D. & Freire, E., The heat capacity of proteins. *Proteins* **22**, 404 (1995).
162. Prabhu, N. V. & Sharp, K. A., Heat capacity in proteins. *Annu. Rev. Phys. Chem.* **56**, 521 (2005).
163. Idiyatullin, D., Nesmelova, I., Daragan, V. A. & Mayo, K. H., Heat capacities and a snapshot of the energy landscape in protein GB1 from the pre-denaturation temperature dependence of backbone NH nanosecond fluctuations. *J. Mol. Biol.* **325**, 149 (2003).
164. Lee, S. Y., Fujitsuka, Y., Kim, D. H. & Takada, S., Roles of physical interactions in determining protein-folding mechanisms: molecular simulation of protein G and alpha spectrin SH3. *Proteins* **55**, 128 (2004).
165. Dumont, C., Emilsson, T. & Gruebele, M., Reaching the protein folding speed limit with large, sub-microsecond pressure jumps. *Nat. Methods* **6**, 515 (2009).
166. Imamura, H., Isogai, Y. & Kato, M., Differences in the structural stability and cooperativity between monomeric variants of natural and de novo Cro proteins revealed by high-pressure Fourier transform infrared spectroscopy. *Biochemistry* **51**, 3539 (2012).
167. Imamura, H. & Kato, M., Effect of pressure on helix-coil transition of an alanine-based peptide: an FTIR study. *Proteins* **75**, 911 (2009).
168. Imamura, H., Isogai, Y., Takekiyo, T. & Kato, M., Effect of pressure on the secondary structure of coiled coil peptide GCN4-p1. *Biochim. Biophys. Acta* **1804**, 193 (2010).

169. Mondragón, A., Subbiah, S., Almo, S. C., Drottar, M. & Harrison, S. C., Structure of the amino-terminal domain of phage 434 repressor at 2.0 Å resolution. *J. Mol. Biol.* **205**, 189 (1989).
170. Laurents, D. V., Corrales, S., Elías-Arnanz, M., Sevilla, P., Rico, M. & Padmanabhan, S., Folding kinetics of phage 434 cro protein. *Biochemistry* **39**, 13963 (2000).
171. Neri, D., Billeter, M., Wider, G. & Wüthrich, K., NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* **257**, 1559 (1992).
172. Clarke, N. D., Kissinger, C. R., Desjarlais, J., Gilliland, G. L. & Pabo, C. O., Structural studies of the engrailed homeodomain. *Protein Sci.* **3**, 1779 (1994).
173. Mayor, U., Johnson, C. M., Daggett, V. & Fersht, A. R., Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13518 (2000).
174. Leahy, D. J., Hendrickson, W. A., Aukhil, I. & Erickson, H. P., Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* **258**, 987 (1992).
175. Clarke, J., Hamill, S. J. & Johnson, C. M., Folding and stability of a fibronectin type III domain of human tenascin. *J. Mol. Biol.* **270**, 771 (1997).
176. Hamill, S. J., Steward, A. & Clarke, J., The folding of an immunoglobulin-like greek key protein is defined by a common-core nucleus and regions constrained by topology. *J. Mol. Biol.* **297**, 165 (2000).
177. Paci, E., Clarke, J., Steward, A., Vendruscolo, M. & Karplus, M., Self-consistent determination of the transition state for protein folding: application to a fibronectin type III domain. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 394 (2003).

178. Feng, S., Kapoor, T. M., Shirai, F., Combs, A. P. & Schreiber, S. L., Molecular basis for the binding of SH3 ligands with non-peptide elements identified by combinatorial synthesis. *Chem. Biol.* **3**, 661 (1996).
179. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M., Amyloid fibril formation by an SH3 domain. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4224 (1998).
180. Grantcharova, V. P. & Baker, D., Folding dynamics of the src SH3 domain. *Biochemistry* **36**, 15685 (1997).
181. Bondos, S. E., Sligar, S. & Jonas, J., High-pressure denaturation of apomyoglobin. *Biochim. Biophys. Acta* **1480**, 353 (2000).
182. Kitahara, R., Yokoyama, S. & Akasaka, K., NMR snapshots of a fluctuating protein structure: ubiquitin at 30 bar-3 kbar. *J. Mol. Biol.* **347**, 277 (2005).
183. Zhai, Y. & Winter, R., Effect of molecular crowding on the temperature-pressure stability diagram of ribonuclease A. *Chem. Phys. Chem.* **14**, 386 (2013).
184. Kalbitzer, H. R., Görler, A., Li, H., Dubovskii, P. V., Hengstenberg, W., Kowolik, C., Yamada, H. & Akasaka, K., ¹⁵N and ¹H NMR study of histidine containing protein HPr from staphylococcus carnosus at high pressure. *Protein Sci.* **9**, 693 (2000).
185. Refaee, M., Tezuka, T., Akasaka, K. & Williamson, M. P., Pressure-dependent changes in the solution structure of hen egg-white lysozyme. *J. Mol. Biol.* **327**, 857 (2003).
186. Hynes, T. R. & Fox, R. O., The crystal structure of staphylococcal nuclease refined at 1.7 Å resolution. *Proteins* **10**, 92 (1991).
187. Lamy, A. & Smith, J., Denaturation of truncated staphylococcal nuclease in molecular dynamics simulation at 300 K. *J. Am. Chem. Soc.* **118**, 7326 (1996).

188. Paliwal, A., Asthagiri, D., Bossev, D. P. & Paulaitis, M. E., Pressure denaturation of staphylococcal nuclease studied by neutron small-angle scattering and molecular simulation. *Biophys. J.* **87**, 3479 (2004).
189. Mitra, L., Smolin, N., Ravindra, R., Royer, C. & Winter, R., Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution—experiments and theoretical interpretation. *Phys. Chem. Chem. Phys.* **8**, 1249 (2006).
190. Hilser, V. J. & Freire, E., Predicting the equilibrium protein folding pathway: structure-based analysis of staphylococcal nuclease. *Proteins* **27**, 171 (1997).
191. Walkenhorst, W. F., Green, S. M. & Roder, H., Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. *Biochemistry* **36**, 5795 (1997).
192. Seemann, H., Winter, R. & Royer, C. A., Volume, expansivity and isothermal compressibility changes associated with temperature and pressure unfolding of staphylococcal nuclease. *J. Mol. Biol.* **307**, 1091 (2001).
193. Roche, J., Dellarole, M., Caro, J. A., Norberto, D. R., Garcia, A. E., Garcia-Moreno, B., Roumestand, C. & Royer, C. A., Effect of internal cavities on folding rates and routes revealed by real-time pressure-jump NMR spectroscopy. *J. Am. Chem. Soc.* **135**, 14610 (2013).
194. Wang, S., Tate, M. W. & Gruner, S. M., Protein crowding impedes pressure-induced unfolding of staphylococcal nuclease. *Biochim. Biophys. Acta* **1820**, 957 (2012).
195. Mishchenko, M. I., Hovenier, J. W. & Travis, L. D., *Light scattering by nonspherical particles: theory, measurements, and applications* (Academic Press, San Diego, 2000).
196. Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A. &

-
- Smith, L. J., Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424 (1999).
197. Lassalle, M. W., Yamada, H. & Akasaka, K., The pressure-temperature free energy-landscape of staphylococcal nuclease monitored by ^1H NMR. *J. Mol. Biol.* **298**, 293 (2000).
198. Privalov, P. L., Cold denaturation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **25**, 281 (1990).

